

Hybrid Type-Logical Grammars, First-Order Linear Logic and the Descriptive Inadequacy of Lambda Grammars*

Richard Moot

May 27, 2014

1 Introduction

Hybrid type-logical grammars (Kubota & Levine 2012, Kubota & Levine 2013*c*, Kubota & Levine 2013*a*) are a relatively new framework in computational linguistics, which combines insights from the Lambek calculus (Lambek 1958) and lambda grammars (Oehrle 1994, Muskens 2001, Muskens 2003) — lambda grammars are also called, depending on the authors, abstract categorial grammars (de Groote 2001) and linear grammars (Pollard 2011), though with somewhat different notational conventions¹. The resulting combined system solves some know problems of both the Lambek calculus and of lambda grammars and the additional expressiveness of hybrid type-logical grammars permits the treatment of linguistic phenomena such as gapping which have no satisfactory solution in either subsystem.

The goal of this paper is to prove that hybrid type-logical grammars are a fragment of first-order linear logic. This embedding result has several important consequences: it not only provides a simple new proof theory for the calculus, thereby clarifying the proof-theoretic foundations of hybrid type-logical grammars, but, since the translation is simple and direct, it also provides several new parsing strategies for hybrid type-logical grammars. Second, NP-completeness of hybrid type-logical grammars follows immediately.

The main embedding result also sheds new light on problems with lambda grammars, which are a subsystem of hybrid type-logical grammars and hence a special case of the translation into first-order linear logic. Abstract categorial grammars are attractive both because of their simplicity — they use the simply

*This work has benefitted from the generous support of the French agency Agence Nationale de la Recherche as part of the project Polymnie (ANR-12-CORD-0004).

¹I prefer the term lambda grammars, since I think it most clearly describes the system. Though the term abstract categorial grammars appears to be more common, and I use it from time to time in this article, I will argue in Section 7 that abstract categorial grammars/lambda grammars are *unlike* all other versions of categorial grammars in important ways.

typed lambda calculus, one of the most widely used tools in formal semantics, to compute surface structure (strings) as well as to compute logical form (meanings) — and because of the fact that they provide a natural account of quantifier scope and extraction; for both, the analysis is superior to the Lambek calculus analysis. So it is easy to get the impression that lambda grammars are an unequivocal improvement over Lambek grammars.

In reality, the picture is much more nuanced: while lambda grammars have some often discussed advantages over Lambek grammars, there are several cases — notably coordination, but we will see in Section 7 that this is true for any analysis where the Lambek calculus uses non-atomic arguments — where the Lambek grammar analysis is clearly superior. Many key examples illustrating the elegance of categorial grammars with respect to the syntax-semantics interface fail to have a satisfactory treatment in abstract categorial grammars.

However, whether or not lambda grammars are an improvement over the Lambek calculus is ultimately not the most important question. Since there is a large number of formal systems which improve upon the Lambek calculus, it makes much more sense to compare lambda grammars to these extensions, which include, among many others, Hybrid Type-Logical Grammars, the Displacement calculus (Morrill, Valentín & Fadda 2011) and multimodal type-logical grammars (Moortgat 1996*b*, Moortgat 1997). These extended Lambek calculi all keep the things that worked in the Lambek calculus but improve on the analysis in ways which allow the treatment of more complex phenomena in syntax and especially in the syntax-semantics interface. Compared to these systems, the inadequacies of lambda grammars are evident: even for the things lambda grammars do right (quantifier scope and extraction), there are phenomena, such as reflexives and gapping, which are handled by the same mechanisms as quantifier scope and extraction in alternative theories, yet which cannot be adequately handled by lambda grammars. The abstract categorial grammar treatment suffers from problems of overgeneration and problems at the syntax-semantics interface unlike any other categorial grammar. I will discuss some possible solutions for lambda grammars, but it is clear that a major redesign of the theory is necessary. The most painless solution seems to be a move either to hybrid type-logical grammars or directly to first-order linear logic: both are simple, conservative extensions which solve the many problems of lambda grammars while staying close to the spirit of lambda grammars.

This paper is structured as follows. Section 2 will introduce first-order linear logic and Section 3 will provide some background about the simply typed lambda calculus. These two introductory sections can be skimmed by people familiar with first-order linear logic and the simply typed lambda calculus respectively. Section 4 will introduce hybrid type-logical grammars and in Section 5 we will give a translation of hybrid type-logical grammars into first-order linear logic and prove its correctness. Section 6 will then compare the Lambek calculus and several of its extensions through their translations in first-order linear logic. This comparison points to a number of potential problems for lambda grammars. We will discuss these problems, as well as some potential solutions in Section 7. Finally, the last section will contain some concluding remarks.

2 First-order Linear Logic

Linear logic was introduced by Girard (1987) as a logic which restricts the structural rules which apply freely in classical logic. The multiplicative, intuitionistic fragment of first-order linear logic (which in the following, I will call either MILL1 or simply first-order linear logic), can be seen as a resource-conscious version of first-order intuitionistic logic. Linear implication, written $A \multimap B$, is a variant of intuitionistic implication $A \Rightarrow B$ with the additional constraint that the A argument formula is used *exactly once*. So, looking at linear logic from the context of grammatical analysis, we would assign an intransitive verb the formula $np \multimap s$, indicating it is a formula which combines with a single np (noun phrase) to form an s (a sentence).

Linear logic is a commutative logic. In the context of language modelling, this means our languages are closed under permutations of the input string, which does not make for a good linguistic principle (at least not a good *universal* one and a principle which is at least debatable even in languages which allow relatively free word order). We need some way to restrict or control commutativity. The Lambek calculus (Lambek 1958) has the simplest such restriction: we drop the structural rule of commutativity altogether. This means linear implication $A \multimap B$ splits into two implications: $A \backslash B$, which looks for an A to its left to form a B , and B / A , which looks for an A to its right to form a B . In the Lambek calculus, we would therefore refine the assignment to intransitive verbs from $np \multimap s$ to $np \backslash s$, indicating the intransitive verb is looking for the subject to its left.

In first-order linear logic, we can choose a more versatile solution, namely using first-order variables to encode word order. We assign atomic formulas a pair of string positions: np becomes $np(0, 1)$, meaning it is a noun phrase spanning position 0 (its leftmost position) to 1 (its rightmost position). Using pairs of (integer) variables to represent strings is standard in parsing algorithms. The addition of quantifiers makes things more interesting. For example, we can assign the formula $\forall x. n(3, x) \multimap np(2, x)$ to a determiner “the” which spans positions 2, 3. This means it is looking for a noun which starts at its right (that is the leftmost position of this noun is the rightmost position of the determiner, 3) but ends at any position x to produce a noun phrase which starts at position 2 (the leftmost position of the determiner) and ends at position x (the rightmost position of the noun). Combined with a noun $n(3, 4)$, this would allow us to instantiate x to 4 and produce $np(2, 4)$. In other words, the formula given to the determiner indicates it is looking for a noun to its right in order to produce a noun phrase, using a form of “concatenation by instantiation of variables” which should be familiar to anyone who has done some logic programming or who has a basic familiarity with parsing in general (Pereira & Shieber 1987, Shieber, Schabes & Pereira 1995). Similarly, we can assign an intransitive verb at position 1, 2 the formula $\forall y. np(y, 1) \multimap s(y, 2)$ to indicate it is looking for a noun phrase to its left to form a sentence, as the Lambek calculus formula $np \backslash s$ for intransitive verbs does — this correspondence between first-order linear logic and the Lambek calculus is fully general and discussed fully in (Moot &

Piazza 2001) and briefly in the next section.

2.1 MILL1

After this informal introduction to first-order linear logic, it is time to be a bit more precise. We will not need function symbols in the current paper, so *terms* are either variables denoted x, y, z, \dots (a countably infinite number) or constants, for which I will normally use integers $0, 1, \dots$, giving an m -word string $m + 1$ string positions, from 0 to m . The atomic formulas are of the form $a(t_1, \dots, t_m)$ with t_i terms, a a predicate symbol (we only need a finite, typically small, number of predicate symbols, often only the following four: n for noun, np for noun phrase, s for sentence, pp for predicate phrase) and m its arity. Our language does *not* contain the identity relation symbol “ $=$ ”. Given this set of atomic formulas \mathcal{A} and the set of variables \mathcal{V} , the set of formulas is defined as follows².

$$\mathcal{F} ::= \mathcal{A} \mid \mathcal{F} \multimap \mathcal{V} \mid \forall \mathcal{V}. \mathcal{F}$$

We treat formulas as syntactically equivalent up to renaming of bound variables, so substituting $\forall y. A[x := y]$ (where A does not contain y before this substitution is made) for $\forall x. A$ inside a formula B will produce an equivalent formula, for example $\forall x. a(x) \equiv \forall y. a(y)$.

Table 1 shows the natural deduction rules for first-order linear logic. The variable x in the $\forall E$ and $\forall I$ rules is called the *eigenvariable* of the rule. The $\forall I$ rule has the condition that the variable y which is replaced by the eigenvariable does not occur in undischarged hypotheses of the proof and that x does not occur in A before the substitution is made³. Throughout this paper, we will use the standard convention in first-order (linear) logic (Girard 1991, Bellin & van de Wiele 1995, Troelstra & Schwichtenberg 2000) that every occurrence of a quantifier \forall, \exists in a sequent uses a distinct variable and in addition that no variable occurs both free and bound in a sequent.

As shown in (Moot & Piazza 2001), we can translate Lambek calculus sequents and formulas into first-order linear logic as follows.

$$\frac{A_1, \dots, A_n \vdash B}{\|A_1\|^{0,1}, \dots, \|A_n\|^{n-1,n} \vdash \|B\|^{0,n}}$$

²We need neither the multiplicative conjunction \otimes nor the existential quantifier \exists in this paper, though adding them to the logic poses no problems. The natural deduction rules for \exists and \otimes are slightly more complicated than those for \forall and \multimap but the basic proof net building blocks don’t change, see for example (Girard 1991, Moot & Piazza 2001, Moot 2014).

³It is sometimes more convenient to use the following $\forall I$ rule

$$\frac{A}{\forall x. A} \quad \forall I^*$$

with the condition there are no free occurrence of x in open hypotheses. The rule of Table 1 is more convenient in the following section when we use meta-variables, where it becomes “replace all occurrences of a (meta-)variable by x , then quantify over x ”.

$$\begin{array}{c}
\frac{A \quad A \multimap B}{B} \multimap E \qquad \frac{[A]^i \quad \vdots \quad B}{A \multimap B} \multimap I \\
\\
\frac{\forall x.A}{A[x := t]} \forall E \qquad \frac{A[y := x]}{\forall x.A} \forall I^*
\end{array}$$

Table 1: Natural deduction rules for first-order linear logic

$$\begin{aligned}
& \|a\|^{x,y} = a(x, y) \\
& \|A/B\|^{x,y} = \forall z. \|B\|^{y,z} \multimap \|A\|^{x,z} \\
& \|B \setminus A\|^{y,z} = \forall x. \|B\|^{x,y} \multimap \|A\|^{x,z}
\end{aligned}$$

The integers 0 to n represent the positions of the formulas in the sequent and the translations for complex formulas introduce universally quantified variables. The translation for A/B states that if we have a formula A/B at positions x, y then for any z if we find a formula B at positions y, z (that is, to the immediate right of our A/B formula) then we have an A at positions x, z , starting at the left position of the A/B formula and ending at the right position of the B argument. In other words, a formula A/B is something which combines with a B to its right to form an A , just like its Lambek calculus counterpart.

Using this translation, we can see that the first-order linear logic formulas used for the determiner and the intransitive verb in the previous section correspond to the translations of np/n at position 2,3 and $np \setminus s$ at position 1,2 respectively.

To give a simple example of a first-order linear logic proof, we shown a derivation of “every student ran”, corresponding to the Lambek calculus sequent.

$$(s/(np \setminus s))/n, n, np \setminus s \vdash s$$

We first translate the sequent into first-order linear logic.

$$\|(s/(np \setminus s))/n\|^{0,1}, \|n\|^{1,2}, \|np \setminus s\|^{2,3} \vdash \|s\|^{0,3}$$

Then translate the formulas as follows.

$$\forall y. [n(1, y) \multimap \forall z. [\forall x. [np(x, y) \multimap s(x, z)] \multimap s(0, z)], n(1, 2), \forall v. [np(v, 2) \multimap s(v, 3) \vdash s(0, 3)]$$

We can then show that “every student ran” is a grammatical sentence under these formula assignments as follows.

$$\begin{array}{c}
\frac{\forall y.[n(1, y) \multimap \forall z.[\forall x.[np(x, y) \multimap s(x, z)] \multimap s(0, z)]]}{n(1, 2) \multimap \forall z.[\forall x.[np(x, 2) \multimap s(x, z)] \multimap s(0, z)]} \forall E \quad n(1, 2) \multimap E \\
\frac{\frac{\forall z.[\forall x.[np(x, 2) \multimap s(x, z)] \multimap s(0, z)]}{\forall x.[np(x, 2) \multimap s(x, 3)] \multimap s(0, 3)} \forall E \quad \forall v.[np(v, 2) \multimap s(v, 3)]}{s(0, 3)} \multimap E
\end{array}$$

The application of the final $\multimap E$ rule is valid, since $\forall x.[np(x, 2) \multimap s(x, 3)] \equiv \forall v.[np(v, 2) \multimap s(v, 3)]$.

Definition 2.1 (Universal closure) If A is a formula we denote the set of free variables of A by $FV(A)$.

For an antecedent $\Gamma = A_1, \dots, A_n$, $FV(\Gamma) = FV(A_1) \cup \dots \cup FV(A_n)$.

The universal closure of a formula A with $FV(A) = \{x_1, \dots, x_n\}$, denoted $Cl(A)$, is the formula $\forall x_1 \dots \forall x_n.A$.

The universal closure of a formula A modulo antecedent Γ , written $Cl_\Gamma(A)$, is defined by universally quantifying over the free variables in A which do not occur in Γ . If $FV(A) \setminus FV(\Gamma) = \{x_1, \dots, x_n\}$, then $Cl_\Gamma(A) = \forall x_1 \dots \forall x_n.A$.

Proposition 2.2 $\Gamma \vdash A$ iff $\Gamma \vdash Cl_\Gamma(A)$.

Proof If the closure modulo Γ prefixes n universal quantifiers to A , we can go from $\Gamma \vdash A$ to $\Gamma \vdash Cl_\Gamma(A)$ by using the $\forall I$ rule n times (the quantified variables added for the closure have been chosen to respect the condition on the rule) and in the opposite direction by using the $\forall E$ rule n times. \square

2.2 MILL1 with focusing and unification

The $\forall E$ rule, as formulated in the previous section, has the disadvantage that it requires us to choose a term t with which to replace x and that making the right choice for t requires some insight into how the resulting formula will be used in the rest of the proof. In the example of the preceding section we need to make two such “educated guesses”: we instantiate y to 2 to allow the elimination rule with minor premiss $n(1, 2)$ and we instantiate z to 3 to produce the desired conclusion $s(0, 3)$.

The standard solution to automate this process in first-order logic theorem proving is to change the $\forall E$ rule: instead of directly replacing the quantified variable by the “right” choice, we replace it by a meta-variable (I will use the Prolog-like notation A, B, \dots for these variables, or, when confusion with the notation A and B for arbitrary formulas is possible $C, D, E, \dots, V, W, X, \dots$). These meta-variables will represent our current knowledge about the term with which we will replace a given quantified variable. The MGU we compute for the endsequent will correspond to the most general instantiations of these variables in the given proof (that is, all other instantiations can be obtained from this final MGU by means of additional substitutions).

The \multimap E rule *unifies* the B formulas of the argument and minor premiss of the rule (so the two occurrences of B need only be unifiable instead of identical). Remember that the unification of two atomic formulas $a(x_1, \dots, x_m)$ and $b(y_1, \dots, y_n)$ is only defined when $a = b$ and $m = n$ and that unification tries to find the most general instantiation of all free variables such that $x_i = y_i$ (for all $1 \leq i \leq n = m$) and fails if no such instantiation exists. The presence of an explicit quantifier presents a complication, but only a minor one: bound variables are treated just like constants which, in addition, must respect the variable condition.

More precisely, the unification of two formulas is defined as follows.

$$\begin{aligned} \text{unify}(a(x_1, \dots, x_n), a(y_1, \dots, y_n)) &= \text{unify}(x_i, y_i) \text{ for all } 1 \leq i \leq n \\ \text{unify}(A_1 \multimap B_1, A_2 \multimap B_2) &= \text{unify}(A_2, A_1), \text{unify}(B_1, B_2) \\ \text{unify}(\forall x.A, \forall y.B) &= \text{unify}(A, B[y := x]) \end{aligned}$$

The \forall case assumes there are no free occurrences of x in B before substitution. It is defined in such a way that it is independent of the actual variable names used for the quantifier (as mentioned, we use a different variable for each occurrence of a quantifier) and bound occurrences of x_i and y_i are treated as constants in the $\text{unify}(x_i, y_i)$ clause, subject to the following condition: if we compute a substitution $D := x$ for a formula A and x is not free for D in A then unification *fails*. In other words, the substitution cannot introduce new bound variables, so for example $\forall y.a(D, y)$ and $\forall z.a(z, z)$ fail to unify, since D is not free for y in $\forall y.a(D, y)$, and therefore we cannot legally substitute y for D since it would result in an “accidental capture”, creating a new bound occurrence of y .⁴

As second problem with natural deduction proof search is that we can have subproofs like the following.

$$\frac{\frac{a(y)}{\forall x.a(x)} \forall I \quad \frac{[A]^i \quad \vdots \quad B}{A \multimap B} \multimap I_i \quad A}{a(y)} \forall E \quad \frac{A \multimap B}{B} \multimap E$$

In both cases, we introduce a connective and then immediately eliminate it. A natural deduction proof is called *normal* if it does not contain any subproof of the forms shown above. One of the classic results for natural deduction is *normalization* which states that we can eliminate such detours (Girard, Lafont & Taylor 1988, Troelstra & Schwichtenberg 2000). In the case of linear logic, removing such detours is even guaranteed to decrease the size of the proof.

We use a form of *focalized* natural deduction (Andreoli 1992, Brock-Nannestad & Schürmann 2010), which is a syntactic variant of natural deduction guaranteed to generate only normal natural deduction proofs. We use two turnstiles,

⁴In such cases, substitution succeeds but does nothing and subsequent unification fails, since the formulas are not alphabetic variants after substitution.

Lexicon

$$A \vdash_n A$$

Axiom/Hypothesis

$$A \vdash_n A$$

Shift Focus

$$\frac{\Gamma \vdash_n A}{\Gamma \vdash_p A} \pm$$

Logical Rules

$$\frac{\Gamma \vdash_n B \multimap A \quad \Delta \vdash_p B}{s(\Gamma), s(\Delta) \vdash_n s(A)} \multimap E$$

$$\frac{\Gamma, B \vdash_p A}{\Gamma \vdash_p B \multimap A} \multimap I$$

$$\frac{\Gamma \vdash_n \forall x.A}{\Gamma \vdash_n A[x := D]} \forall E$$

$$\frac{\Gamma \vdash_p A[y := x]}{\Gamma \vdash_p \forall x.A} \forall I^*$$

Table 2: Focused first-order linear logic with unification

the negative \vdash_n and the positive \vdash_p (for the reader familiar with focused proofs, $\Gamma \vdash C \Downarrow$ corresponds to $\Gamma \vdash_n C$ and $\Gamma \vdash C \Uparrow$ to $\Gamma \vdash_p C$).

We will call a sequent $\Gamma \vdash_p C$ a *positive* sequent (and C a positive formula) and a sequent $\Gamma \vdash_n C$ a *negative* sequent (and C a negative formula).

Table 2 shows the rules of first-order linear logic in this format. For the lexicon rule, we require that the formula A is closed. The formula A of the hypothesis rule *can* contain free variables.

For the $\forall I$ rule, y is either a variable or a meta-variable which has no free occurrences in any undischarged hypothesis.

For the $\multimap E$ rule, s is the most general unifier of $\langle \Gamma, B \rangle$ and $\langle \Delta, B \rangle$. That is, we unify the two occurrences of B in their respective contexts, using unification for complex formulas as defined above. The resulting most general unifier is then applied to the two contexts and to A (replacing, if necessary, any variables

shared between A and B in the formula A).

We can see from the rules that axioms start negative and stay negative as long as they are the major premiss of a $\neg E$ rule or the premiss of a $\forall E$ rule. We must switch to positive sequents to use the introduction rules or to use the sequent as the minor premiss of a $\neg E$ rule.

The “detour” subproofs we have seen above cannot receive a consistent labeling: the formula $A \neg B$ is the conclusion of a $\neg I$ rule and must therefore be on the right-hand side of a positive sequent, however, it is also the major premiss of a $\neg E$ rule and must therefore be on the right-hand side of a negative sequent (it is easily verified there is no way to transform a positive sequent into a negative sequent, however the point is that the *original* detour receives an inconsistent labeling).

$$\begin{array}{c}
 \frac{\frac{\frac{\vdash_p a(y)}{\vdash_p \forall x.a(x)} \forall I}{\vdash_n \forall x.a(x)} \forall E \quad \frac{\frac{\frac{[\vdash_n A]^i}{\vdash_p B} \neg I_i}{\vdash_p A \neg B} \neg I_i}{\vdash_n A \neg B} \neg E \quad \frac{\vdash_p A}{\vdash_n B} \neg E}{\vdash_n a(y)} \forall E
 \end{array}$$

Definition 2.3 A principal branch is a sequence of negative sequents which starts at a hypothesis, then follows all elimination rules from (major) premiss to conclusion ending at a focus shift rule (this corresponds to the normal notion of principal branch from e.g. (Girard et al. 1988); a sequence of negative sequents can only pass through the major premiss of a $\neg E$ rule and through the single premiss of a $\forall E$ rule).

A track is a path of negative sequents followed by a focus shift followed by a path of positive sequents. A track ends either in the conclusion of the proof or in the minor premiss of a $\neg E$ rule.

The main track of a proof is the track which ends in its conclusion (these definitions corresponds to the standard notion of track and main track in normal proofs, see e.g. (Troelstra & Schwichtenberg 2000)).

This suggests a relation between focused proofs and normal natural deduction proofs, which is made explicit in the following two propositions.

Proposition 2.4 For every natural deduction proof of $\Gamma \vdash B$, there is a focused natural deduction proof with unification of $\Gamma \vdash_p B$.

Proof We first transform the natural deduction proof of $\Gamma \vdash B$ into a normal natural deduction proof, then proceed by induction on the length of the proof and show that we can create both a proof of $\Gamma \vdash_p B$ and a substitution s . We proceed by induction on the depth of the proof.

If $d = 1$, we have an axiom or hypothesis rule, which we translate as follows.

$$\frac{A \vdash_n A}{A \vdash_p A} \pm$$

If $d > 1$ we proceed by case analysis on the last rule.

The only case which requires some attention is the $\multimap E$ case. Given that the proof is normal, we have a normal (sub)proof which ends in a $\multimap E$ rule. We are therefore on the principal branch of this subproof and we know that a principal branch starts with an axiom/lexicon rule then passes only $\forall E$ rules and $\multimap E$ rules through their major premiss. Hence, the last rule producing the major premiss in the original proof must either have been an axiom/lexicon rule or an elimination rule for \multimap or \forall .

Now induction hypothesis gives us a proof δ_1 of $\Gamma \vdash_p B \multimap A$ and a proof δ_2 of $\Delta \vdash_p B$. However, given that the last rule of the proof which produces δ_1 was either axiom/lexicon, the $\forall E$ rule or the $\multimap E$ rule — all of which have negative sequents as their conclusion — the last rule of δ_1 must have been the focus shift rule. Removing this focus shift rule produces a valid proof $\delta_{1'}$ of $\Gamma \vdash_n B \multimap A$, which we can combine with the proof δ_2 of $\Delta \vdash_p B$ as follows.

$$\frac{\begin{array}{c} \vdots \delta_{1'} \\ \Gamma \vdash_n B \multimap A \end{array} \quad \begin{array}{c} \vdots \delta_2 \\ \Delta \vdash_p B \end{array}}{\frac{\Gamma, \Delta \vdash_n A}{\Gamma, \Delta \vdash_p A} \pm}$$

Note that this is again a proof which ends with a focus shift rule.

Since the original proof uses the stricter notion of *identity* (instead of unifiability) for the B formulas, we need not change the substitution we have computed so far and therefore leave Γ , Δ and A unchanged.

For the $\forall E$ rule, induction hypothesis gives us a proof δ of $\Gamma \vdash_p \forall x.A$, by reasoning similar to the case for $\multimap E$, we know the last rule of δ was a focus shift rule, which we can remove, then extend the proof as follows.

$$\frac{\begin{array}{c} \vdots \delta \\ \Gamma \vdash_n \forall x.A \end{array}}{\frac{\Gamma \vdash_n A[x := D]}{\Gamma \vdash_p A[x := D]} \pm} \forall E$$

Adding the substitution $D := t$ (where t is the term used for the in the original $\forall E$ rule) to the unifier.

The cases for $\forall I$ and $\multimap I$ are trivial, since we can extend the proof with the same rule. \square

Proposition 2.5 *For every focused natural deduction proof, there is a natural deduction proof.*

Proof If we remove the focus shift rule and replace both \vdash_n and \vdash_p by \vdash then we only need to give specific instantiations for the $\forall E$ rules. The most general unifier ς computed for the complete proof gives us such values for each (negatively) quantified variable (if wanted, remaining meta-variables can be replaced by free variables). \square

The following is a standard property of normal natural deduction proofs (and therefore of focused natural deduction proofs).

Proposition 2.6 *Focused proofs satisfy the subformula property. That is, any formula occurring in a proof of $\Gamma \vdash_p B$ (or $\Gamma \vdash_n B$) is a subformula either of Γ or of B .*

The following proposition is easily verified by induction on A and using the correspondence between natural deduction proofs and λ -terms.

Proposition 2.7 *We can restrict the focus shift rule to atomic formulas A . When we do so, we only produce long normal form proofs (which correspond to beta normal eta long lambda terms).*

The proof from the previous section looks as follows in the unification-based version of first-order linear logic, though we use a form with implicit antecedents to economize on horizontal space and to make comparison with the proof of the previous section easier. This proof produces the most general unifier $Y = 2$, $Z = 3$, corresponding to the explicit instantiations for y and z at the $\forall E$ rules in the previous proof.

$$\frac{\frac{\frac{\frac{\vdash_n \forall y.[n(1, y) \multimap \forall z.[\forall x.[np(x, y) \multimap s(x, z)] \multimap s(0, z)]]}{\vdash_n n(1, Y) \multimap \forall z.[\forall x.[np(x, Y) \multimap s(x, z)] \multimap s(0, z)]} \forall E \quad \frac{\vdash_n n(1, 2)}{\vdash_p n(1, 2)} \pm}{\vdash_n n(1, 2)} \multimap E \quad \frac{\frac{\frac{\vdash_n \forall z.[\forall x.[np(x, 2) \multimap s(x, z)] \multimap s(0, z)]}{\vdash_n \forall x.[np(x, 2) \multimap s(x, Z)] \multimap s(0, Z)} \forall E \quad \frac{\vdash_n s(0, 3)}{\vdash_p s(0, 3)} \pm}{\vdash_n s(0, 3)} \pm \quad \frac{\frac{\frac{\vdash_p \forall v.[np(v, 2) \multimap s(v, 3)]}{\vdash_p \forall v.[np(v, 2) \multimap s(v, 3)]} \pm}{\vdash_p \forall v.[np(v, 2) \multimap s(v, 3)]} \multimap E \quad \frac{\vdash_n s(0, 3)}{\vdash_p s(0, 3)} \pm}{\vdash_p s(0, 3)} \multimap E$$

Restricting focus shift (\pm) to atomic formulas, produces the following proof in long normal form. Remark that our hypothesis in this proof is not $np(V, 2)$ but $np(U, W)$ which unifies with $np(V, 2)$ at the $\multimap E$ rule immediately below it.

$$\frac{\frac{\frac{\frac{\frac{\vdash_n \forall y.[n(1, y) \multimap \forall z.[\forall x.[np(x, y) \multimap s(x, z)] \multimap s(0, z)]]}{\vdash_n n(1, Y) \multimap \forall z.[\forall x.[np(x, Y) \multimap s(x, z)] \multimap s(0, z)]} \forall E \quad \frac{\vdash_n n(1, 2)}{\vdash_p n(1, 2)} \pm}{\vdash_n n(1, 2)} \multimap E \quad \frac{\frac{\frac{\vdash_n \forall z.[\forall x.[np(x, 2) \multimap s(x, z)] \multimap s(0, z)]}{\vdash_n \forall x.[np(x, 2) \multimap s(x, Z)] \multimap s(0, Z)} \forall E \quad \frac{\vdash_n s(0, 3)}{\vdash_p s(0, 3)} \pm}{\vdash_n s(0, 3)} \pm \quad \frac{\frac{\frac{\frac{\frac{\vdash_n np(U, W)}{\vdash_p np(U, W)} \pm \quad \frac{\vdash_n \forall v.[np(v, 2) \multimap s(v, 3)]}{\vdash_n np(V, 2) \multimap s(V, 3)} \forall E}{\vdash_p np(V, 2) \multimap s(V, 3)} \multimap E \quad \frac{\frac{\vdash_n s(V, 3)}{\vdash_p s(V, 3)} \pm}{\vdash_p np(V, 2) \multimap s(V, 3)} \multimap I_1 \quad \frac{\vdash_p \forall w.[np(w, 2) \multimap s(w, 3)]}{\vdash_p \forall w.[np(w, 2) \multimap s(w, 3)]} \forall I}{\vdash_p \forall w.[np(w, 2) \multimap s(w, 3)]} \multimap E \quad \frac{\vdash_n s(0, 3)}{\vdash_p s(0, 3)} \pm}{\vdash_p s(0, 3)} \multimap E$$

2.3 Proof Nets

Proof nets are an elegant alternative to natural deduction and an important research topic in their own right; for reasons of space we provide only an informal introduction — the reader interested in more detail is referred to (Girard 1995) for an introduction and to (Danos & Regnier 1989, Bellin & van de Wiele 1995) for detailed proofs in the context of linear logic and to (Lamarche & Retoré 1996,

Moot 2002, Moot & Retoré 2012) for introductions in the context of categorial grammars and the Lambek calculus. Though proof nets shine especially for the \exists and \otimes rules (where the natural deduction formulation requires commutative conversions to decide proof equivalence), they are a useful alternative in the \forall and \multimap case as well since they provide an easy combinatorial way to do proof search and therefore make arguments about non-derivability of statements and serve to count the number of readings.

Girard (1991) shows that the proof nets of multiplicative linear logic (Girard 1987, Danos & Regnier 1989) have a simple extension to the first-order case. Essentially, a proof net is a graph labeled with (polarized occurrences of) the (sub)formulas of a sequent $\Gamma \vdash C$, subject to some conditions we will discuss below. Obviously, not all graphs labeled with formulas correspond to derivable statements. However, we can characterize the proof nets among the larger class of proof structures (graphs labeled with formulas which, contrary to proof nets, do *not* necessarily correspond to proofs) by means of simple graph-theoretic properties.

The basic building blocks of proof structures are *links*, as shown in Figure 1. We will call the formulas displayed below the link their *conclusion* and the formulas displayed above it their premisses. The axiom link (top left) has no premisses and two conclusions, the cut link has no conclusions and two premisses, the binary logical links have two premisses (A and B) and one conclusion $A \multimap B$ and the unary logical links have one premiss A and one conclusion $\forall x.A$. We will call x the eigenvariable of the link and require that all links use distinct variables.

Given a statement $A_1, \dots, A_n \vdash C$ we can unfold the formulas using the logical links of the figure, using the negative links for the A_i and the positive link for C . Since there is only one type of link for each combination of connective/polarity, we unfold our formulas deterministically⁵, until we end up at the atomic formulas and have produced a “formula forest”, a sequence of formula decomposition trees labeled with some additional information (polarity labels and dashed lines), which is sometimes called a *proof frame*.

We turn this proof frame into a *proof structure* by connecting atomic formulas of opposite polarity in such a way there is a perfect matching between the positive and negative atoms. This step can already fail, for example if the number of positive and negative occurrences of an atomic formula differ but also because of incompatible atomic formulas like $a(0, 1)$ and $a(x, 1)$, with x the eigenvariable of a \forall^+ link. More generally, it can be the case that there is no coherent substitution which allows us to perform a complete matching of the atomic formulas using axiom links. These restrictions on the instantiations of variables are a powerful tool for proof search (Moot 2007, Moot 2014).

Proof structures are essentially graphs where some of the links are drawn with dashed lines; the binary dashed lines are paired, as indicated by the connecting arc. We will call the dashed logical links (\forall^+ and \multimap^+) the *positive* links

⁵For the negative \forall this is not immediately obvious, since we need to choose a suitable term t . We will discuss this case below but we will essentially use meta-variables and unification just like we did for natural deduction in Section 2.2.

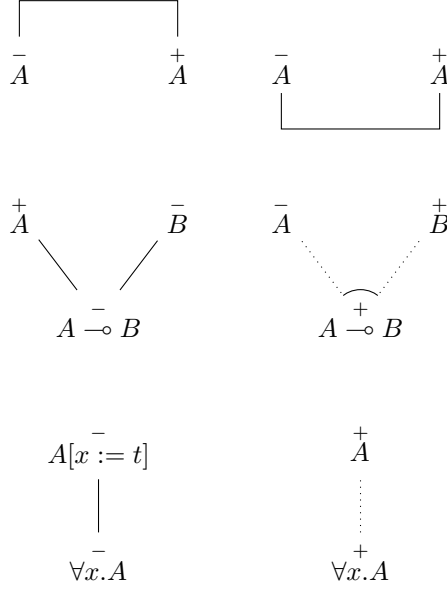


Figure 1: Links for proof structures in the $\forall, -\circ$ fragment of first-order linear logic.

and the solid logical links (\forall^- and $-\circ^-$) the *negative* links. The terms positive and negative links only apply to the logical links; the axiom and cut link are neither positive nor negative. A proof structure containing only negative logical links is just a graph labeled with polarized formulas.

Figure 2 shows the proof net which corresponds to the natural deduction proof of Section 2.1. To save space, we have noted only the main connective at each link; the full formula can be obtained unambiguously from the context. We have also been free in the way we ordered the premisses of the $-\circ$ links, which allows us to give a planar presentation of the axiom links, much like Lambek calculus proof nets. However, there is no planarity requirement in the proof net calculus; the first-order variables offer more flexibility than simple planarity. For the \forall^- links, we have annotated the substitutions next to the link. If we use a unification-based presentation, as we did for natural deduction in Section 2.2, we can “read off” these substitutions from the most general unifier computed for the axioms (as opposed to natural deduction, the axioms and not the $-\circ E$ rule, which corresponds to the $-\circ^-$ link, are responsible for the unification of variables).

A proof structure is a *proof net* if the statement $A_1, \dots, A_n \vdash C$ is derivable, that is, given the proof of Section 2.1, we know the proof structure of Figure 2 is a proof net. However, this definition is not very useful, since it depends on finding a proof in some other proof system; we would like to use the proof structure

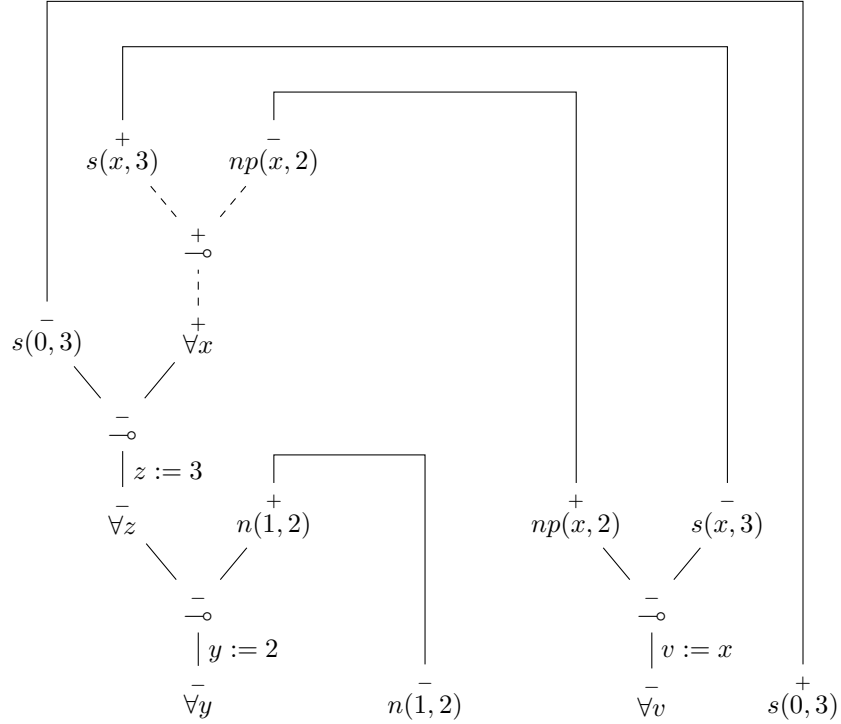


Figure 2: Proof net corresponding to the natural deduction proof of Section 2.1

itself to directly decide whether or not the statement is derivable. However, it is possible to distinguish the proof nets from the other proof structures by simple graph-theoretic properties. To do so, we first introduce some auxiliary notions, which turn the graph-like proof structures into standard graphs. Since axiom, cut and the negative links already produce normal graphs (\multimap^- corresponds to two edges, all other links to a single edge in the graph), we only need a way to remove the positive links.

Definition 2.8 *A switching is a choice for each positive link as follows.*

- For each \multimap^+ link, we choose one its premisses (A or B).
- For each \forall^+ link, we choose either its premiss A or any of the formulas in the proof structure containing a free occurrence of the eigenvariable of the link.

A given a switching s , a correction graph is a proof structure where we replace all dashed links by a link from the conclusion of the link to the formula chosen by the switching s .

Theorem 2.9 (*Girard 1991*) *A proof structure is a proof net iff all its correction graphs are acyclic and connected.*

Defined like this, it would seem that deciding whether or not a proof structure is a proof net is rather complicated: there are potentially many correction graphs — we have two independent possibilities for each \multimap^+ link and generally at least two subformulas containing the eigenvariable of each \forall^+ link, giving 2^n correction graphs for n positive links — and we need verify all of them. Fortunately, there are very efficient alternatives: linear time in the quantifier-free case (Murawski & Ong 2000, Guerrini 1999) and at most squared time, though possibly better, in the case with quantifiers (Moot 2014).

Going back to the example shown in Figure 2, we can see that there are two positive links and twelve correction graphs: there are six free occurrences of x — four in atomic formulas and two additional occurrences in the conclusions (\multimap^+ and \multimap^-) which combine these atomic formulas into $np(x, 2) \multimap s(x, 2)$ — times the two independent possibilities for switching \multimap^+ left or right. We can verify that all twelve possibilities produce acyclic, connected graphs. Removing the positive links splits the graph into three connected components: the single node labeled \multimap^+ (representing $(np(x, 2) \multimap s(x, 2))^+$), a component containing the intransitive verb ending at the axioms to $s(x, 3)^+$ and $np(x, 2)^-$ and a final component containing the rest of the graph, ending at the conclusion of the \forall^+ link (which has been disconnected from its premiss). Now, any switching for the \multimap^+ link will connect its isolated conclusion node to the component containing $s(x, 3)^+$ and $np(x, 2)^-$ (via one or the other of these nodes), leaving two connected components. Finally, all free occurrences of the variable x occur in this newly created component, therefore any choice for a switching of the \forall^+ link will join these disconnected components into a single, connected component. Since each choice connected two disjoint components, we have not generated any cycles.

We can also show that this is the only possible proof structure for the given logical statement: there is only one choice for the n formulas, one choice for the np formulas though two choices for the s formulas. However, the alternative proof structure would link $s(0, z)$ to $s(x, z)$ (for some value of z), which fails because x , being the eigenvariable of a \forall^+ link, cannot be instantiated to 0.

As a second example, let's show how we can use correction graphs to show undervivability. Though it is clear that the switching for the universal quantifier must refer to free occurrences of its eigenvariable somewhere (as do its counterparts in natural deduction and sequent calculus), it is not so easy to find a small example in the \forall, \multimap fragment where this condition is necessary to show undervivability, since finding a global instantiation of the variables is already a powerful constraint on proof structures. However, the existential quantifier and the universal quantifier differ only in the labeling of formulas for the links and we need the formula labeling only for determining the free variables.

A proof structure of the undervivable sequent $(\forall x.a(x)) \multimap b \not\vdash \exists y.[a(y) \multimap b]$ is shown in Figure 3. It is easy to verify this is the unique proof structure corresponding to this sequent. This sequent is used for computing the prenex

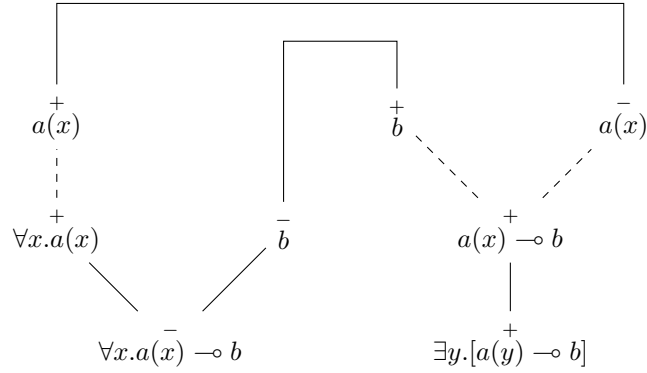


Figure 3: Proof structure which is not a proof net

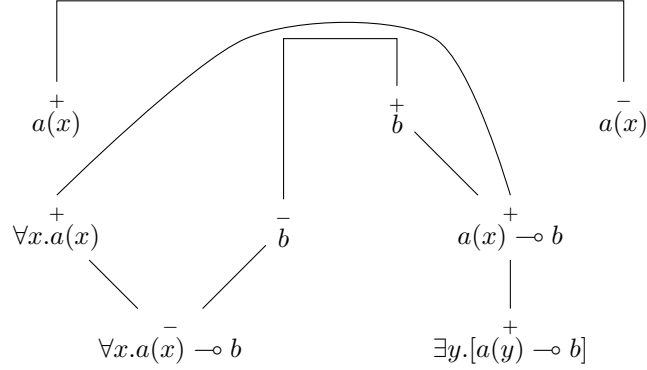


Figure 4: A cyclic and disconnected correction graph for the proof structure of Figure 3

normal form of a formula in classical logic (replacing \multimap by \Rightarrow), but it is invalid in intuitionistic logic and linear logic since it depends on the structural rule of right contraction.

In order to show the sequent is invalid in linear logic, it suffices to find a switching such that the corresponding correction graph either contains a cycle or is disconnected. Figure 4 shows a correction graph for the proof structure of Figure 3 which is both cyclic and disconnected: the axiom $a(x) \vdash a(x)$ is not connected to the rest of the structure and the connection between $\forall x.a(x)$ and $a(x) \multimap b$ produces a cycle, since there is a second path to these two formulas through the axiom $b \vdash b$.

This concludes our brief introduction to proof nets for first-order linear logic. We refer the reader to Appendix A of (Girard et al. 1988) for discussion about the relation between proof nets and natural deduction.

$$\begin{array}{c}
x^\alpha \vdash x : \alpha \\
\\
\frac{\Gamma \vdash M : \alpha \rightarrow \beta \quad \Delta \vdash N : \alpha}{\Gamma, \Delta \vdash (M N) : \beta} \rightarrow E \\
\\
\frac{\Gamma, x^\alpha \vdash M : \beta}{\Gamma \vdash \lambda x. M : \alpha \rightarrow \beta} \rightarrow I
\end{array}$$

Table 3: Curry-style typing rules for the linear lambda calculus

3 Basic Properties of the Simply Typed Lambda Calculus

Before introducing hybrid type-logical grammars, we will first review some basic properties of the simply typed lambda calculus which will prove useful in what follows. This section is not intended as a general introduction to the simply typed lambda calculus: we will assume the reader has at least some basic knowledge such as can be found in Chapter 3 of (Girard et al. 1988) or other textbooks and some knowledge about substitution and most general unifiers. For more detail, and for proofs of the lemmas and propositions of this section, the reader is referred to (Hindley 2008).

A remark on notation: we will use \rightarrow exclusively as a type constructor (also when we know we are using it to type a linear lambda term) and \multimap exclusively as a logical connective.

Definition 3.1 *A lambda term M is a linear lambda term iff*

1. *for every subterm $\lambda x. N$ of M , x has exactly one occurrence in N (in other words, each abstraction binds exactly one variable occurrence),*
2. *all free variables of M occur exactly once.*

Table 3 lists the Curry-style typing rules for the linear lambda calculus. For the $\rightarrow E$ rule, Γ and Δ cannot share term variables; for the $\rightarrow I$ rule, Γ cannot contain x (ie. Γ, x^α must be a valid context).

Proposition 3.2 *For linear lambda terms, we have the following:*

1. *When M is a linear lambda term and $\Gamma \vdash M : \alpha$ a deduction of M , then the variables occurring in Γ are exactly the free variables of M .*
2. *If M, N are linear lambda terms which do not share free variables then $(M N)$ is a linear lambda term.*
3. *If M is a linear lambda term with a free occurrence of x then $\lambda x. M$ is a linear lambda term.*

4. If M is a linear lambda term and $M \rightarrow_{\beta\eta} N$ then N is a linear lambda term.

Lemma 3.3 (Substitution) *If $\Gamma, x : \alpha \vdash M : \beta$, $\Delta \vdash N : \alpha$ and Γ and Δ are compatible (ie. there are no conflicting variable assignments and therefore Γ, Δ is a valid context), then $\Gamma, \Delta \vdash M[x := N] : \beta$.*

The following two results are rather standard, we can find them in (Hindley 2008) as Lemmas 2C1 and 2C2.

Lemma 3.4 (Subject Reduction) *Let $M \rightarrow_{\beta\eta} N$, then $\Gamma \vdash M : \alpha \Rightarrow \Gamma \vdash N : \alpha$*

Lemma 3.5 (Subject Expansion) *Let $M \rightarrow_{\beta\eta} N$ with M a linear lambda term, then $\Gamma \vdash N : \alpha \Rightarrow \Gamma \vdash M : \alpha$*

3.1 Principal types

The main notions from Chapter 3 of (Hindley 2008) are the following.

Definition 3.6 (Principal type) *A principal type of a term M is a type α such that*

1. *for some context Γ we have $\Gamma \vdash M : \alpha$*
2. *if $\Gamma' \vdash M : \beta$, then there is a substitution s such that $s(\alpha) = \beta$.*

Definition 3.7 (Principal pair) *A principal pair for a term M is a pair $\langle \Gamma, \alpha \rangle$ such that $\Gamma \vdash M : \alpha$ and for all β such that $\Gamma \vdash M : \beta$ there is a substitution s with $s(\alpha) = \beta$*

Definition 3.8 (Principal deduction) *A principal deduction for a term M is a derivation δ of a statement $\Gamma \vdash M : \alpha$ such that every other derivation with term M is an instance of δ (ie. obtained by globally applying a substitution s to all types in the proof).*

From the definitions above, it is clear that if δ is a principal deduction for $\Gamma \vdash M : \alpha$ then Γ, α is a principal pair and α a principal type of M .

If M contains free variables x_1, \dots, x_n we can compute the principal type $\alpha_1 \rightarrow \dots (\alpha_n \rightarrow \beta)$ of the closed term $\vdash \lambda x_1, \dots, x_n. M$ which is the same as the principal type for $x_1^{\alpha_1}, \dots, x_n^{\alpha_n} \vdash M^\beta$.

3.2 The principal type algorithm

The principal type algorithm of Hindley (2008) is defined as follows. It is slightly more general and computes principal *deductions*. It takes as input a lambda term M and outputs either its principal type α or fails in case M is untypable. We closely follow Hindley's presentation, keeping his numbering but restricting ourselves to linear lambda terms; we omit his correctness proof of the algorithm.

We proceed by induction on the construction of M .

- I. If M is a variable, say x , then we take an unused type variable α and return $x^\alpha \vdash x : \alpha$ as principal deduction.
- II. If M is of the form $\lambda x.N$ and x occurs in N then we look at the principal deduction δ of N by induction hypothesis : if we fail to compute a principal deduction for N then there is no principal deduction for $\lambda x.N$ either. If such a deduction δ does exist, then we can extend it as follows.

$$\frac{\begin{array}{c} \vdots \delta \\ x^\alpha, \Gamma \vdash N : \beta \end{array}}{\Gamma \vdash \lambda x.N : \alpha \rightarrow \beta} \multimap I$$

- III. M is of the form $\lambda x.N$ and x does not occur in N ; this case cannot occur since it violates the condition on linear lambda terms (we must bind exactly one occurrence of x in N), so we fail.
- IV. M is of the form (NP) . If the algorithm fails for either N or P , then M is untypable and we fail. If not, induction hypothesis gives us a principal proof δ_1 for $\Gamma \vdash N : \gamma'$ and a principal proof δ_2 for $\Delta \vdash P : \gamma$. If necessary, we rename type variables if Γ and Δ such that Γ and Δ have no type variables in common. Since M is linear, N and P cannot share term variables.

- (a) If γ' is of the form $\alpha \rightarrow \beta$ then we compute the most general unifier \mathfrak{s} of $\langle \Gamma, \alpha \rangle$ and $\langle \Delta, \gamma \rangle$. If this fails the term is untypable; if not we combine the proofs as follows.

$$\frac{\begin{array}{c} \vdots \mathfrak{s}(\delta_1) \\ \mathfrak{s}(\Gamma) \vdash N : \mathfrak{s}(\alpha) \rightarrow \mathfrak{s}(\beta) \end{array} \quad \begin{array}{c} \vdots \mathfrak{s}(\delta_2) \\ \mathfrak{s}(\Delta) \vdash P : \mathfrak{s}(\gamma) \end{array}}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash (NP) : \mathfrak{s}(\beta)} \rightarrow E$$

- (b) If γ' is a type variable, then we compute the most general unifier \mathfrak{s} of $\langle \Gamma, \gamma' \rangle$ and $\langle \Delta, \gamma \rightarrow \beta \rangle$ (with β a fresh type variable). If this succeeds and the term is typable, we can produce its principal proof as follows.

$$\frac{\begin{array}{c} \vdots \mathfrak{s}(\delta_1) \\ \mathfrak{s}(\Gamma) \vdash N : \mathfrak{s}(\gamma) \rightarrow \mathfrak{s}(\beta) \end{array} \quad \begin{array}{c} \vdots \mathfrak{s}(\delta_2) \\ \mathfrak{s}(\Delta) \vdash P : \mathfrak{s}(\gamma) \end{array}}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash (NP) : \mathfrak{s}(\beta)} \rightarrow E$$

The main utility of principal types in the current paper is given by the coherence theorem.

Theorem 3.9 (Coherence) *Suppose $\Gamma \vdash N : \alpha$ and let α be a principal type of N then $\forall P \forall \Gamma' \subseteq \Gamma \quad \Gamma' \vdash P : \alpha \implies P \equiv_{\beta\eta} N$*

The coherence theorem states that a principal type determines a lambda term uniquely (up to $\beta\eta$ equivalence). Since we work in a linear system, where weakening is not allowed, we only need the special case $\Gamma' = \Gamma$. This special case of Theorem 3.9 is the following: if $\Gamma \vdash N : \alpha$ with α a principal type of N then for any P such that $\Gamma \vdash P : \alpha$ we have that $P \equiv_{\beta\eta} N$.

In brief, the principal type algorithm allows us to compute the principal type of a given typable lambda term, whereas the coherence theorem allows us to reconstruct a lambda term (up to $\beta\eta$ equivalence) from a principal type.

Definition 3.10 *We say a sequent $\Gamma \vdash C$ is balanced if all atomic types occurring in the sequent occur exactly twice.*

The following lemmas are easy consequences of 1) the Curry-Howard isomorphism between linear lambda terms and Intuitionistic Linear Logic (ILL), which allows us to interpret the linear type constructor “ \multimap ” as the logical connective “ \multimap ” 2) the correspondence between (normal) natural deduction proofs and (cut-free) proof nets and 3) the fact that renaming the conclusions of the axiom links in a proof net gives another proof net.

Lemma 3.11 *If M is a linear lambda term with free variables x_1, \dots, x_n then the principal type $\alpha_1 \multimap \dots (\alpha_1 \multimap \beta)$ of $\lambda x_1 \dots \lambda x_n. M$ is balanced. Hence the principal type of $x_1^{\alpha_1}, \dots, x_n^{\alpha_n} \vdash M^\beta$ is balanced.*

Proof Compute the natural deduction proof of M and convert it to a ILL proof net. By subject reduction (Lemma 3.4), normalization/cut elimination keeps the type α invariant. Let P be the cut-free proof net which corresponds to the natural deduction proof of M and which has the same type as M . We obtain a balanced proof net by using a different atomic formula for all axiom links. From this proof net, we can obtain all other types of M by renaming the axiom links (allowing for non-atomic axiom links), hence it is a principal type and it is balanced by construction. \square

Lemma 3.12 *If M is a beta-normal lambda term with free variables x_1, \dots, x_n and if $\lambda x_1, \dots, \lambda x_n. M$ has a balanced typing then M is linear.*

Proof If $\lambda x_1, \dots, \lambda x_n. M$ has a balanced typing, then from this typing we can construct a unique cut-free ILL proof net of $\lambda x_1, \dots, \lambda x_n. M$. Since it is an ILL proof net, this lambda term must be linear and therefore M as well. \square

3.3 Examples

To illustrate the principal type algorithm, we give two examples in this section.

As a first example, we compute the principal proof of $\mathbf{C} \equiv \lambda f. \lambda x. \lambda y. ((f y) x)$ as follows.

$$\begin{array}{c}
\frac{f^{\gamma_0} \vdash f : \gamma_0 \quad y^\beta \vdash y : \beta}{f^{\beta \rightarrow \gamma_1}, y^\beta \vdash (f y) : \gamma_1} \rightarrow E \quad x^\alpha \vdash x : \alpha \\
\hline
\frac{f^{\beta \rightarrow \alpha \rightarrow \gamma}, y^\beta, x^\alpha \vdash ((f y) x) : \gamma}{f^{\beta \rightarrow \alpha \rightarrow \gamma}, x^\alpha \vdash \lambda y.((f y) x) : \beta \rightarrow \gamma} \rightarrow I \\
\hline
\frac{f^{\beta \rightarrow \alpha \rightarrow \gamma}, x^\alpha \vdash \lambda y.((f y) x) : \beta \rightarrow \gamma}{f^{\beta \rightarrow \alpha \rightarrow \gamma} \vdash \lambda x. \lambda y.((f y) x) : \alpha \rightarrow \beta \rightarrow \gamma} \rightarrow I \\
\hline
\vdash \lambda f. \lambda x. \lambda y.((f y) x) : (\beta \rightarrow \alpha \rightarrow \gamma) \rightarrow \alpha \rightarrow \beta \rightarrow \gamma \rightarrow I
\end{array}$$

The substitutions $\gamma_0 := \beta \rightarrow \gamma_1$ (for the topmost $\rightarrow E$ rule) and $\gamma_1 := \alpha \rightarrow \gamma$ (for the bottom $\rightarrow E$ rule) have been left implicit in the proof.

As a second example, the principal proof of $l^{2 \rightarrow 1} \vdash \lambda O. \lambda S. \lambda z. (S (l (O z)))$ is the following.

$$\begin{array}{c}
\frac{O^{\alpha_0} \vdash O : \alpha_0 \quad z^\beta \vdash z : \beta}{O^{\beta \rightarrow \alpha_1}, z^\beta \vdash (O z) : \alpha_1} \rightarrow E \\
\frac{l^{2 \rightarrow 1} \vdash l : 2 \rightarrow 1 \quad O^{\beta \rightarrow \alpha_1}, z^\beta \vdash (O z) : \alpha_1}{l^{2 \rightarrow 1}, O^{\beta \rightarrow 2}, z^\beta \vdash (l (O z)) : 1} \rightarrow E \\
\frac{S^{\alpha_2} \vdash S : \alpha_2 \quad l^{2 \rightarrow 1}, O^{\beta \rightarrow 2}, z^\beta \vdash (l (O z)) : 1}{S^{1 \rightarrow \alpha}, l^{2 \rightarrow 1}, O^{\beta \rightarrow 2}, z^\beta \vdash (S (l (O z))) : \alpha} \rightarrow E \\
\frac{S^{1 \rightarrow \alpha}, l^{2 \rightarrow 1}, O^{\beta \rightarrow 2}, z^\beta \vdash (S (l (O z))) : \alpha}{S^{1 \rightarrow \alpha}, l^{2 \rightarrow 1}, O^{\beta \rightarrow 2} \vdash \lambda z. (S (l (O z))) : \beta \rightarrow \alpha} \rightarrow I \\
\frac{l^{2 \rightarrow 1}, O^{\beta \rightarrow 2} \vdash \lambda z. (S (l (O z))) : \beta \rightarrow \alpha}{l^{2 \rightarrow 1}, O^{\beta \rightarrow 2} \vdash \lambda S. \lambda z. (S (l (O z))) : (1 \rightarrow \alpha) \rightarrow \beta \rightarrow \alpha} \rightarrow I \\
\hline
l^{2 \rightarrow 1} \vdash \lambda O. \lambda S. \lambda z. (S (l (O z))) : (\beta \rightarrow 2) \rightarrow (1 \rightarrow \alpha) \rightarrow \beta \rightarrow \alpha \rightarrow I
\end{array}$$

The substitutions $\alpha_0 := \beta \rightarrow \alpha_1$, $\alpha_1 := 2$, $\alpha_2 := 1 \rightarrow \alpha$ (of the three $\rightarrow E$ rules, from top to bottom) have again been left implicit.

4 Hybrid Type-Logical Grammars

Hybrid type-logical grammars have been introduced in (Kubota & Levine 2012) as an extension of lambda grammars which combines insights from the Lambek calculus into lambda grammars. Depending on authors, lambda grammars (Muskens 2003) are also called abstract categorial grammars (de Groote 2001) or linear grammars (Pollard 2011).

Formulas of hybrid type-logical grammars are defined as follows, where \mathcal{F}_2 are the formulas of hybrid type-logical grammars and \mathcal{F}_1 the formulas of Lambek grammars. \mathcal{A} denotes the atomic formulas of the Lambek calculus — we will call these formulas *simple* atomic formulas, since their denotations are strings — \mathcal{B} signifies complex atomic formulas, whose denotations are not simple strings, but string *tuples*.

$$\begin{aligned}
\mathcal{F}_0 &::= \mathcal{A} \\
\mathcal{F}_1 &::= \mathcal{F}_0 \mid \mathcal{F}_1 / \mathcal{F}_1 \mid \mathcal{F}_1 \setminus \mathcal{F}_1 \\
\mathcal{F}_2 &::= \mathcal{B} \mid \mathcal{F}_1 \mid \mathcal{F}_2 | \mathcal{F}_2
\end{aligned}$$

As is clear from the recursive definition of formulas above, hybrid type-logical grammars are a sort of layered or fibred logic. Such logics have been studied before as extensions of the Lambek calculus by replacing the atomic formulas in \mathcal{F}_0 by feature logic formulas (Bayer & Johnson 1995, Dörre & Manandhar 1995).

Lambek grammars are obtained by not allowing connectives or complex atoms in \mathcal{F}_2 . From hybrid type-logical grammars, we obtain lambda grammars by not allowing connectives in \mathcal{F}_1 . Inversely, we can see hybrid type-logical grammars as lambda grammars where simple atomic formulas have been replaced by Lambek formulas.

Before presenting the rules of hybrid type-logical grammars, we'll introduce some notational conventions: A and B range over arbitrary formulas; C , D and E denote type variables or type constants; n and $n - 1$ denote type constants corresponding to string positions; α and β denote arbitrary types. Types are written as superscripts to the terms; x , y and z denote term variables; M and N denote arbitrary terms.

Table 4 shows the rules of Hybrid Type-Logical Grammars. The rules are presented in such a way that they compute principal types in addition to the terms. We obtain the Church-typed version — equivalent to the calculus presented in (Kubota & Levine 2012) — by replacing all type variables and constants by the type constant σ . For the principal types, we use the Curry-typed version, though for readability, we often write the types of subterms as superscripts as well.

The subsystem containing only the rules for $|$ is simply lambda grammar. The subsystem containing only the rules for $/$ and \backslash is a notational variant of the Lambek calculus.

For the Lexicon rule, $\langle x^{n \rightarrow n-1}, \alpha \rangle$ is a principal pair for M or, equivalently, with $\lambda x.M$ a β -normal η -long linear lambda term and $(n \rightarrow n - 1) \rightarrow \alpha$ its principal type). For the Axiom/Hypothesis rule, M is the eta-expansion of $x : A$.

For the Lambek calculus elimination rule $/E$ and $\backslash E$, s is the most general unifier of $\langle \Gamma; F \rangle$ and $\langle \Delta; D \rangle$ (this generally just replaces F by D but takes care of the cases where $C = D$ or $E = F$ as well). The concatenation operation of the Lambek calculus corresponds to function composition on terms and to unification of string positions on types (much like we have seen in Section 2).

For the Lambek calculus introduction rules $/I$ and $\backslash I$, s is the most general unifier of $\langle \Gamma; C \rightarrow D \rangle$ (resp. $\langle \Gamma; D \rightarrow C \rangle$) and $\langle \emptyset; F \rightarrow F \rangle$ (ie. we simply identify C and D and replace x by the identity function on string positions — the empty string).

In the $|E$ rule, s is the most general unifier of $\langle \Gamma; \beta \rangle$ and $\langle \Delta; \gamma \rangle$.

For convenience, we will often tacitly apply the following rule.

$$\frac{\Gamma \vdash M^\alpha : A \quad M =_{\beta\eta} N}{\Gamma \vdash N^\alpha : A} =_{\beta\eta}$$

Though the above rule is not strictly necessary, we use it to simplify the lambda terms we compute, performing on-the-fly β -normalization (ie. we replace

Lexicon

$$x^{n \rightarrow n-1} : A \vdash M^\alpha : A$$

Axiom/Hypothesis

$$\overline{x^\alpha : A \vdash M^\alpha : A}$$

Logical rules – Lambek

$$\frac{\Gamma \vdash M^{F \rightarrow C} : A/B \quad \Delta \vdash N^{E \rightarrow D} : B}{s(\Gamma), s(\Delta) \vdash (\lambda z^{s(E)}. M (N z))^{s(E) \rightarrow s(C)} : A} /E$$

$$\frac{\Gamma \vdash M^{F \rightarrow C} : B \quad \Delta \vdash N^{E \rightarrow D} : B \setminus A}{s(\Gamma), s(\Delta) \vdash (\lambda z^{s(E)}. M (N z))^{s(E) \rightarrow s(C)} : A} \setminus E$$

$$\frac{\Gamma, x^{D \rightarrow C} : B \vdash M^{D \rightarrow E} : A}{s(\Gamma) \vdash ((\lambda x^{s(D) \rightarrow s(C)}. M) (\lambda z^{s(F)}. z))^{s(C) \rightarrow s(E)} : A/B} /I$$

$$\frac{\Gamma, x^{C \rightarrow D} : B \vdash M^{E \rightarrow D} : A}{s(\Gamma) \vdash ((\lambda x^{s(C) \rightarrow s(D)}. M) (\lambda z^{s(F)}. z))^{s(E) \rightarrow s(C)} : B \setminus A} \setminus I$$

Logical rules – lambda grammars

$$\frac{\Gamma \vdash M^{\beta \rightarrow \alpha} : A|B \quad \Delta \vdash N^\gamma : B}{s(\Gamma), s(\Delta) \vdash (M N)^{s(\alpha)} : A} |E$$

$$\frac{\Gamma, x^\beta : B \vdash M^\alpha : A}{\Gamma \vdash (\lambda x^\beta. M^\alpha)^{\beta \rightarrow \alpha} : A|B} |I$$

Table 4: Logical rules for hybrid type-logical grammars

M by its beta-normal, or beta-normal-eta-long, form N). Since we have both subject reduction and subject expansion, M and N are guaranteed to have the same type α .

Apart from the types, the system presented in Table 4 is a notational variant of hybrid type-logical grammars as presented by Kubota and Levine (2012, 2013c). We have replaced strings as basic types by string *positions* with Church type $\sigma \rightarrow \sigma$. This is a standard strategy in abstract categorial grammars, akin to the difference lists in Prolog, which allows us to do without an explicit concatenation operation: concatenation is simply treated as function

composition, as can be seen from the term assignments for the $/E$ and $\backslash E$ rules. The introduction rules $/I$ and $\backslash I$ are presented somewhat differently than the Kubota and Levine version, who present rules requiring (in our notation) premisses with term assignments $M \equiv_{\beta\eta} \lambda z.N[(xz)]$ and $M \equiv_{\beta\eta} (x N)$ respectively. The present formulation has the advantage that it is more robust in the sense that it does not require us to test that M is $\beta\eta$ equivalent to the given terms. Though it may appear a bit strange that the $/I$ and $\backslash I$ rules require the identity of the type variable D between x and M , it is clear that this follows from the intended interpretation, which requires the string variable x to occur at the beginning (resp. end) of the string denoted by M , and this solution seems preferable to interleaving normalization and pattern matching in our rules.

The types, at least for the $|$ rules, are exactly those computed using the principal type algorithm of Hindley (2008) discussed in Section 3.1. We will see how the types for the Lambek connectives and the lexicon rule correspond to principal type computations in the next section.

4.1 Justification of the principal types for the new rules

For $/E$ and $\backslash E$, their principal types are justified as follows; \mathfrak{s}_1 is the most general unifier of $\langle z^G; G \rangle$ and $\langle \Delta; E \rangle$ — since G is a type variable not occurring elsewhere, we can assume without loss of generality that \mathfrak{s}_1 just replaces G with E — and \mathfrak{s}_2 is the most general unifier of $\langle \mathfrak{s}_1(\Delta), \mathfrak{s}_1(z^G); \mathfrak{s}_1(D) \rangle$ and $\langle \Gamma; F \rangle$. The important type unification is of D and F (the unification of E and G affects only a discharged axiom).

At the level of the types, the two rules are the same: both correspond to concatenation.

$$\frac{\frac{\frac{z^G \vdash z : G \quad \Delta \vdash N : E \rightarrow D}{\mathfrak{s}_1(\Delta), \mathfrak{s}_1(z^G) \vdash N z : \mathfrak{s}_1(D)} \rightarrow E \quad \Gamma \vdash M : F \rightarrow C}{\mathfrak{s}_2(\Gamma), \mathfrak{s}_2(\mathfrak{s}_1(\Delta)), \mathfrak{s}_2(\mathfrak{s}_1(z^G)) \vdash M(N z) : \mathfrak{s}_2(\mathfrak{s}_1(C))} \rightarrow E}{\mathfrak{s}_2(\Gamma), \mathfrak{s}_2(\mathfrak{s}_1(\Delta)) \vdash \lambda z.M(N z) : \mathfrak{s}_2(\mathfrak{s}_1(E)) \rightarrow \mathfrak{s}_2(\mathfrak{s}_1(C))} \rightarrow I$$

Taking $\mathfrak{s} = \mathfrak{s}_1 \cup \mathfrak{s}_2$, which is possible since Γ , Δ and z are disjoint, gives us the following proof.

$$\frac{\frac{\frac{z^G \vdash z : G \quad \Delta \vdash N : E \rightarrow D}{\mathfrak{s}(\Delta), \mathfrak{s}(z^G) \vdash N z : \mathfrak{s}(D)} \rightarrow E \quad \Gamma \vdash M : F \rightarrow C}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta), \mathfrak{s}(z^G) \vdash M(N z) : \mathfrak{s}(C)} \rightarrow E}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash \lambda z.M(N z) : \mathfrak{s}(E) \rightarrow \mathfrak{s}(C)} \rightarrow I$$

Since \mathfrak{s}_1 only replaced G by E and G no longer appears in the conclusion of the proof (the corresponding hypothesis z has been withdrawn) we can treat \mathfrak{s} as the most general unifier of $\langle \Delta; D \rangle$ and $\langle \Gamma; F \rangle$.

We compute the principal type for the $/I$ rule as follows.

$$\frac{\frac{\Gamma, x^{D \rightarrow C} \vdash M : D \rightarrow E}{\Gamma \vdash \lambda x.M : (D \rightarrow C) \rightarrow D \rightarrow E} \rightarrow I \quad \frac{\overline{z^F \vdash z : F}}{\vdash \lambda z.z : F \rightarrow F} \rightarrow I}{\Gamma \vdash (\lambda x.M)\lambda z.z : C \rightarrow E} \rightarrow E$$

And symmetrically for $\backslash I$.

$$\frac{\frac{\Gamma, x^{C \rightarrow D} \vdash M : E \rightarrow D}{\Gamma \vdash \lambda x.M : (C \rightarrow D) \rightarrow E \rightarrow D} \rightarrow I \quad \frac{\overline{z^F \vdash z : F}}{\vdash \lambda z.z : F \rightarrow F} \rightarrow I}{\Gamma \vdash (\lambda x.M)\lambda z.z : E \rightarrow C} \rightarrow E$$

From the point of view of the principal type computation, we identify the C and D variables, essentially replacing x by the empty string.

Lemma 4.1 *The proof rules for Hybrid type-logical grammars of Table 4 compute principal types for the lambda terms corresponding to their proofs.*

Proof We essentially use the same algorithm as Hindley (2008), which is somewhat simplified by the restriction to linear lambda terms which are eta-long.

The principal types for $/E$, $\backslash E$, $/I$ and $\backslash I$ rules are justified as shown above.

The lexicon rule is justified by the Substitution Lemma (Lemma 3.3): given a principal type α for a lexical entry, we replace a hypothesis of the form $\alpha \vdash \alpha$ by a hypothesis of the form $n \rightarrow n - 1 \vdash \alpha$, where we know this second sequent has a linear proof. \square

Corollary 4.2 *Given a principal type derived by the rules of hybrid type-logical grammar shown above, we can compute the corresponding lambda term up to $\beta\eta$ equivalence.*

Proof Since the principal types computed are balanced by Lemma 3.11, by the Coherence theorem (Theorem 3.9), we can compute the corresponding lambda term up to $\beta\eta$ equivalence. An easy way to do so is to construct the proof net corresponding to the principal type (which is unique because of balance) and to compute its lambda term; this lambda term is the unique beta-normal eta-long term corresponding to the principal type. \square

4.2 Example

As an example of how to compute the principal derivation corresponding to a hybrid derivation, we look at the following hybrid derivation.

$$\frac{\frac{\frac{[x : np] \quad [y : np \backslash s]}{\lambda z.(x(yz)) : s} \backslash E}{\lambda x \lambda z.(x(yz)) : s | np} | I \quad \lambda P \lambda v.((P e) v) : s | (s | np)}{| E} \quad \frac{\lambda v.(e(yv)) : s}{\lambda w.(e w) : s / (np \backslash s)} / I$$

The corresponding principal derivation looks as follows (for reasons of vertical space, the lexical entry for $e^{1 \rightarrow 0}$ has not been eta-expanded to $\lambda P \lambda w.(P e) w$ as it should to obtain the given principal type instead of $((1 \rightarrow 0) \rightarrow G) \rightarrow G$; though either type will end up being instantiated to the same result type, the eta-expanded principal type $((1 \rightarrow 0) \rightarrow H \rightarrow G) \rightarrow H \rightarrow G$ has the important advantage that it can be obtained without instantiating type variables to complex types; similarly, $x^{A \rightarrow B}$ and $y^{C \rightarrow D}$ appear in eta-short form).

$$\begin{array}{c}
\frac{x^{A \rightarrow B} \vdash x^{A \rightarrow B} \quad \frac{\frac{y^{C \rightarrow D} \vdash y^{C \rightarrow D} \quad \frac{z^E \vdash z^E}{\rightarrow E}}{y^{C \rightarrow D}, z^C \vdash (y z)^D} \rightarrow E}{x^{D \rightarrow B}, y^{C \rightarrow D}, z^C \vdash (x (y z))^B} \rightarrow E \\
\frac{x^{D \rightarrow B}, y^{C \rightarrow D} \vdash (\lambda z.x (y z))^{C \rightarrow B} \rightarrow I}{y^{C \rightarrow D} \vdash (\lambda x.\lambda z.x (y z))^{(D \rightarrow B) \rightarrow C \rightarrow B} \rightarrow I} \rightarrow I \quad \frac{e^{1 \rightarrow 0} \vdash (\lambda P.P e)^{((1 \rightarrow 0) \rightarrow H \rightarrow G) \rightarrow H \rightarrow G}}{e^{1 \rightarrow 0}, y^{C \rightarrow 1} \vdash (\lambda z.e (y z))^{C \rightarrow 0} \rightarrow I} \rightarrow E \\
\frac{e^{1 \rightarrow 0} \vdash (\lambda y.\lambda z.e (y z))^{(C \rightarrow 1) \rightarrow C \rightarrow 0} \rightarrow I \quad \frac{v^J \vdash v^J}{\vdash (\lambda v.v)^{J \rightarrow J} \rightarrow I}}{e^{1 \rightarrow 0} \vdash (\lambda z.e z)^{1 \rightarrow 0}}
\end{array}$$

The $\backslash E$ and the $/I$ rules correspond to three rules each in this principal derivation (the derivation of $\lambda z.x (y z)$ for $\backslash E$ and the part of the derivation from $\lambda z.e (y z)$ to $\lambda z.e z$ for $/I$, this last rule satisfies the constraint for the application of the rule, with y appearing at the last position)

In principle, the computation of the principal type can fail because of the constants (even though there might be a proof using variables). However, this failure would mean the final term fails to respect the word order of the input string. Principal types using distinct variables for string positions would seem a useful tool for computing all possible word orders for a given set of lexical entries, though.

4.3 Semantics

One of the attractive points of categorial grammars is that we have a very simple and elegant syntax-semantics interface by means of the Curry-Howard isomorphism between intuitionistic proofs and lambda terms (or, in our case between linear intuitionistic proofs and linear lambda terms). By interpreting the logical connectives for the implications “/”, “\”, “|” and “ \multimap ” as the type constructor “ \rightarrow ” — the *formulas as types* interpretation — our derivations in the Lambek calculus, in lambda grammars, in hybrid type-logical grammars and in first-order linear logic (where we treat the quantifier as being semantically inert, that is, quantifier rules are “invisible” to the meaning) correspond to λ -terms — the *proofs as terms* interpretation. Using the Curry-Howard isomorphism, we can obtain semantics in the tradition of Montague simply by giving lexical substitutions in the lexicon, using essentially the rules of Table 3 (though we typically use the Church-style typing) to assign a derivational meaning to a proof.

The semantic version of the proof from the previous section looks as follows.

$$\begin{array}{c}
\frac{[x : np] \quad [Q : np \backslash s]}{(Q x) : s} \backslash E \\
\frac{\lambda x.(Q x) : s | np}{\lambda P.\forall z.(P z) : s | (s | np)} | I \\
\frac{\lambda P.\forall z.(P z) : s | (s | np)}{\forall z.(Q z) : s} | E \\
\frac{\forall z.(Q z) : s}{\lambda Q.\forall z.(Q z) : s / (np \backslash s)} / I
\end{array}$$

Though syntactically, the Lambek elimination rule corresponds to function composition (concatenation), *semantically* it corresponds to simple application and the introduction rule to abstraction. Given the standard Montegovian semantics for “everyone” as $\lambda P.\forall z.(P z)$ (the set of properties P such that all z have this property), the previous proof actually produces an equivalent term as the semantics for $s/(np \backslash s)$, so the generalized quantifier can function as a Lambek calculus subject quantifier while keeping the same semantics.

More detail about the syntax-semantics interface in categorial grammars can be found in (Moortgat 1997, Moot & Retoré 2012).

5 Equivalence

For the main result, we only need to show that a hybrid principal type proof corresponds to a MILL1 proof, since we can reconstruct the lambda term from the principal type.

The basic idea which makes the correspondence work is that there is a 1-1 mapping between the atomic terms of a predicate in MILL1 and the principal type which is assigned to the corresponding term in a hybrid derivation. So from the term assigned to a hybrid derivation, we compute the principal type using the principal type algorithm (PTA) and this gives us the first-order variables and from the first-order variables of a MILL1 derivation we obtain the principal type and a hybrid lambda term thanks to the coherence theorem, as shown schematically below.

$$\text{Hybrid lambda term} \xrightleftharpoons[\text{Coherence}]{\text{PTA}} \text{Principal type} \longleftrightarrow \text{First-order variables}$$

5.1 String positions, types and formulas

We need an auxiliary function f (for *flatten*) which reduces a complex type to a list of atomic types. Following Kanazawa (2011), we compute this list by first taking the yield of the type tree and then reversing this list, which is convenient for induction since it has $f(\beta \rightarrow \alpha) = f(\alpha) \frown f(\beta)$ (“ \frown ” denotes list concatenation, $[A]$ the singleton list containing element A and $[A_1, \dots, A_n]$ the n -element list with i th element A_i).

Definition 5.1 Let α be a type, the list $f(\alpha)$ is defined as follows.

$$\begin{aligned} f(A) &= [A] \quad \text{when } A \text{ atomic} \\ f(\beta \rightarrow \alpha) &= f(\alpha) \frown f(\beta) \end{aligned}$$

For example, we have the following.

$$\begin{aligned} &f((B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A) \\ &= f((1 \rightarrow A) \rightarrow B \rightarrow A) \frown f(B \rightarrow 2) \\ &= f(B \rightarrow A) \frown f(1 \rightarrow A) \frown f(B \rightarrow 2) \\ &= [A, B, A, 1, 2, B] \end{aligned}$$

Definition 5.2 Let A be a formula in Hybrid Type-Logical Grammar, α its principal type and $L = f(\alpha)$ the flattened list of atomic types obtained from α according to Definition 5.1. The translation of A into first-order linear logic is defined as follows.

$$\begin{aligned} \|p\|^{[C_1, \dots, C_n]} &= p(C_1, \dots, C_n) \\ \|(A|B)\|^{f(\beta \rightarrow \alpha)} &= \|B\|^{f(\beta)} \multimap \|A\|^{f(\alpha)} \\ \|(A/B)\|^{[C, D]} &= \forall x. \|B\|^{[D, x]} \multimap \|A\|^{[C, x]} \\ \|(B \setminus A)\|^{[C, D]} &= \forall x. \|B\|^{[x, C]} \multimap \|A\|^{[x, D]} \end{aligned}$$

We can obtain a closed formula by universally quantifying over all variables in the list of arguments replacing all of them with quantified variables using the universal closure operation (Definition 2.1).

$$\|A\|_c^L = Cl(\|A\|^L)$$

Proposition 5.3 Let A be a formula in first-order linear logic and H a formula in hybrid type-logical grammar and $A \equiv \|H\|^{f(\alpha)}$. The free meta-variables of A are exactly the type variables of α (and of $f(\alpha)$).

Proof Immediate by induction on H using the translation. All new variables introduced during the translation are bound. \square

Lemma 5.4 Let A_1 and A_2 be first-order linear logic formulas obtained by the translation function from Hybrid Type-Logical Grammar formulas H_1 and H_2 with γ_1 and γ_2 as their respective principal types. In other words, $A_1 \equiv \|H_1\|^{f(\gamma_1)}$ and $A_2 \equiv \|H_2\|^{f(\gamma_2)}$.

A_1 unifies with A_2 with MGU \mathfrak{s} if and only if $H_1 \equiv H_2$ and γ_1 unifies with γ_2 with this same MGU \mathfrak{s} .

Proof Suppose A_1 and A_2 unify with MGU \mathfrak{s} . We must show that $H_1 \equiv H_2$ and that \mathfrak{s} is an MGU for γ_1 and γ_2 . Showing $H_1 \equiv H_2$ is an easy induction (exploiting the fact that $A|B$ does not have a quantifier prefix and therefore cannot unify with a Lambek connective and that A/B and $B \setminus A$ cannot unify

with each other because of the condition preventing accidental capture of variables). Given that H_1 and H_2 are identical, we know that A_1 and A_2 differ only in the free variables (the bound variables are equivalent up to renaming) and that the free variables for A_1 and A_2 are exactly the type variables of γ_1 and γ_2 (by Proposition 5.3). Therefore any substitution that makes A_1 and A_2 equal (up to renaming of bound variables) makes γ_1 and γ_2 equal.

For the other direction, suppose that $H_1 \equiv H_2$ and that s is the MGU of γ_1 and γ_2 . Since s is a MGU $s(\gamma_1) \equiv s(\gamma_2)$ and therefore given that the translation function uses identical hybrid formulas and identical principal types we have that $A_1 \equiv A_2$. \square

It is insightful to compare the translation of $(np \setminus s)/np$ (with principal type $2 \rightarrow 1$) to that of $(s|np)|np$ with principal type $(B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A$. Though the two end results are formulas which are equivalent to each other (after universal closure of the meta-variables), there is a difference in the string position list for the non-atomic subformulas: the Lambek formula only ever has a pair of string positions, whereas the linear formula starts with a full list of string positions which decreases at each step. In other words, for the Lambek formula, we compute the string positions step-by-step whereas the lambda grammar version of the same formula precomputes all string positions then divides them among the subformulas.

$$\begin{aligned}
& \|(np \setminus s)/np\|^{[1,2]} \\
&= \forall y. \|np\|^{[2,y]} \multimap \|np \setminus s\|^{[1,y]} \\
&= \forall y. np(2, y) \multimap \|np \setminus s\|^{[1,y]} \\
&= \forall y. np(2, y) \multimap \forall x. \|np\|^{[x,1]} \multimap \|s\|^{[x,y]} \\
&= \forall y. np(2, y) \multimap \forall x. [np(x, 1) \multimap s(x, y)]
\end{aligned}$$

$$\begin{aligned}
& \|(s|np)|np\|^{[A,B,A,1,2,B]} \\
&= \|np\|^{[2,B]} \multimap \|s|np\|^{[A,B,A,1]} \\
&= np(2, B) \multimap \|s|np\|^{[A,B,A,1]} \\
&= np(2, B) \multimap \|np\|^{[A,1]} \multimap \|s\|^{[A,B]} \\
&= np(2, B) \multimap np(A, 1) \multimap s(A, B)
\end{aligned}$$

$$\begin{aligned}
& \|(s|np)|np\|_c^{[A,B,A,1,2,B]} \\
&= \forall x. \forall y. \{np(2, B) \multimap np(A, 1) \multimap s(A, B)\} [A := x, B := y] \\
&= \forall x. \forall y. [np(2, y) \multimap np(x, 1) \multimap s(x, y)]
\end{aligned}$$

Remember that sequents in hybrid type-logical grammar are of the form $x_1^{\alpha_1} : A_1, \dots, x_n^{\alpha_n} : A_n \vdash M^\beta : B$ with M a linear lambda term containing exactly the free variables x_1, \dots, x_n and that the principal type of $\lambda x_1, \dots, x_n. M$ is balanced and of the form $\alpha_1 \rightarrow \dots \rightarrow \alpha_n \rightarrow \beta$. For the translation, we separate lexical axioms from other axioms: lexical axioms correspond to closed formulas, whereas the other axioms typically have free variables. With this in mind, we translate sequents as $\|A_1\|^{f(\alpha_1)}, \dots, \|A_n\|^{f(\alpha_n)} \vdash \|B\|^{f(\beta)}$, where the translation $\|\cdot\|_c$ is used for hypotheses which start at a lexicon rule and $\|\cdot\|$ for hypothesis which start at the axiom rule. For the right-hand side B , we use

the universal closure of all free variables in $f(\beta)$ *minus* the free variables on the left hand side of the sequent (the only free variables are those used in the translation of hypothesis rules), this is the universal closure of B modulo Γ of Definition 2.1.

In order not to overburden our notation, when the types are understood from the context, we will often abbreviate this translation as $\|\Gamma\| \vdash \|B\|$ (or even as $\Gamma \vdash \|B\|$, leaving the translation of Γ implicit). As a special case of this translation, the sequent $w_1^{1 \rightarrow 0} : A_1, \dots, w_n^{n \rightarrow n-1} : A_n \vdash M^{n \rightarrow 0} : B$, which is the endsequent corresponding to a sentence in a hybrid type-logical grammars, is translated as $\|A_1\|_c^{[0,1]}, \dots, \|A_n\|_c^{[n-1,n]} \vdash \|B\|_c^{[0,n]}$.

Example: gapping To give an example, the gapping lexical entry for “and” of (Kubota & Levine 2012) looks as follows in our notation.

$$((s|tv)|(s|tv)) : \lambda STV2 \lambda STV1 \lambda TV \lambda z. (STV1 \ TV)(and \ (STV2 \ \lambda x.x))$$

where tv is short for $(np \backslash s)/np$. The principal type for this lambda term would be the following (the corresponding formulas have been annotated above for ease of comparison).

$$(\overbrace{(E \rightarrow E)}^{tv} \rightarrow \overbrace{D \rightarrow 4}^s) \rightarrow ((\overbrace{(B \rightarrow A)}^{tv} \rightarrow \overbrace{3 \rightarrow C}^s) \rightarrow (\overbrace{(B \rightarrow A)}^{tv} \rightarrow \overbrace{D \rightarrow C}^s))$$

If tv were an atomic formula, the first-order linear logic formula would look as shown below on the first line, the complete formula (for the positive translation) is shown just below it.

$$\begin{aligned} (tv(E, E) \multimap s(4, D)) \multimap (tv(A, B) \multimap s(C, 3)) \multimap tv(A, B) \multimap s(C, D) \equiv \\ (\forall v.[np(v, E) \multimap \forall w.[np(E, w) \multimap s(v, w)]] \multimap s(4, D)) \multimap \\ (\forall x'[np(x, A) \multimap \forall y'[np(B, y') \multimap s(x', y')]] \multimap s(C, 3)) \multimap \\ \forall x.[np(x, A) \multimap \forall y.[np(B, y) \multimap s(x, y)]] \multimap s(C, D) \end{aligned}$$

Though the formula above looks intimidating (even before universal closure), it is easy to verify that it is *equivalent* (up to variable names) to the first-order linear logic formula which corresponds to the analysis of gapping for the Displacement calculus from Section 3.2.6 of (Morrill et al. 2011), using the translation given in (Moot 2014).

5.2 Proof-theoretic properties of the translation into MILL1

Before proving the main theorem, stating that for every hybrid proof there is a first-order linear logic proof of its translation, we will spend some time on the structure of normal/focused natural deduction proofs and the consequences of the translation function. Given that in hybrid type-logical grammars, the lambda-grammar connective “|” always outscopes the Lambek connectives “/” and “\”, proofs using the translated formulas into focused first-order linear logic look schematically as shown in Figure 5.

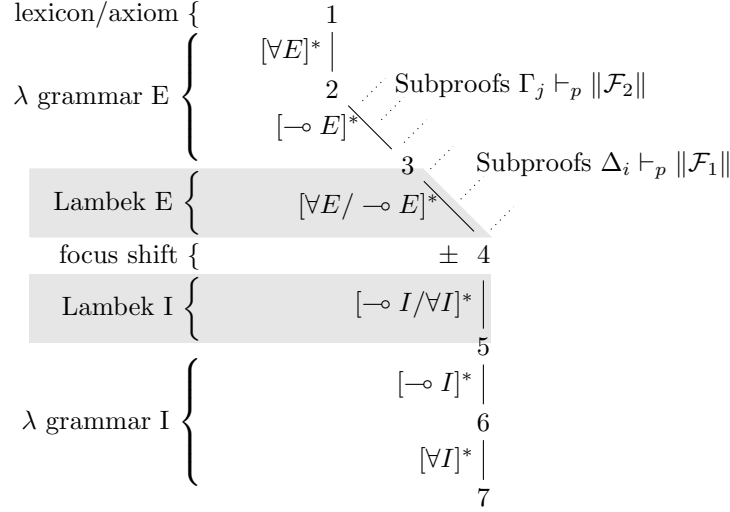


Figure 5: Schematic form of the main track of a translated hybrid sequent.

The figure shows the main track of a proof, which starts either with a hypothesis/axioms, then has an elimination part, followed by a focus shift followed by an introduction part ending in the conclusion of the proof — this is just the definition of a main track (Definition 2.3). The definition of formulas guarantees that the elimination part starts with any number of $[E]$ rules (possibly zero, like all other parts, as indicated by the $*$ superscript in the figure) followed by any combination of $[/E]$ and $[\backslash E]$ rules. The order is inverse in the introduction part of the track, with $[/I]$ and $[\backslash I]$ preceding $[I]$. For the translation of these rules into first-order linear logic, the quantifiers corresponding to the rules for the lambda grammar connective “ \vdash ” are obtained by universal closure, so if they are present, it must be as a prefix at the beginning of the proof or as a postfix at the end of the proof — the subpaths labeled (1)-(2) and (6)-(7) in Figure 5 — and the Lambek connectives correspond to a combination of a \forall and a \multimap rule upon translation.

Proposition 5.5 a. *The main track of a focused proof of a translated hybrid sequent looks as shown in Figure 5.*

b. *The subproofs $\Gamma_j \vdash_p \|\mathcal{F}_2\|$ contain the sequence of proof steps in (1)-(6), that is they do not end with any $\forall I$ rules corresponding to a hybrid connective.*

c. *The subproofs in $\Delta_i \vdash_p \|\mathcal{F}_1\|$ contain the sequence of proof steps in (1)-(5), that is they do not end with any lambda grammar introduction rules.*

Proof These are immediate consequences of the translation function and the structure of normal proofs.

a. follows from the way the translation function is defined and the standard structure of a main track.

b. since normal proofs satisfy the subformula property and since hybrid connectives are translated into prenex formulas, we do not produce subformulas of the form $(\forall x_1, \dots, x_n.[A \multimap B]) \multimap C$ (for $n \geq 1$).

c. would contradict the definition of hybrid formulas, since it would have a Lambek connective outscope a lambda grammar connective. \square

An immediate corollary of Proposition 5.5 is that $\forall I$ rules corresponding to lambda grammar connectives occur only at the end of the main track of a proof, just like $\forall E$ rules corresponding to lambda grammar connective occur only at the start of any track in which they occur.

Definition 5.6 *We say a first-order linear logic proof obtained by translating a hybrid proof is in quantifier-reduced form, when all $\forall E$ and $\forall I$ rules obtained by universal closure of lambda-grammar connectives have been removed from the proof.*

More precisely, the translation is kept as before with the following two exceptions:

- the Lexicon rule is translated as $\|A\|_c^{f(\alpha)} \vdash \|A\|^{f(\alpha)}$ (with the closure operation applied only to the translation of the antecedent)
- the endsequent is translated as $\Gamma \vdash \|C\|^{f(\beta)}$ (without the usual closure modulo Γ).

Proposition 5.7 *A sequent $\Gamma \vdash A$ produced by the translation function is derivable if and only if its quantifier-reduced form is.*

Proof Immediate by Proposition 2.2. \square

Quantifier-reduced form is a way of “compiling” away the predictable prefixes of $\forall E$ rules (for each of the lexical leaves of the proof) and the equally predictable postfix of $\forall I$ rules introduced by the universal closure operation. This simplifies the structure of the proof, as is clear from Proposition 5.8 below and from Figure 5 — we keep only the subpath (2)-(6). It also simplifies the correctness proof of the translation in the following sections, since we avoid having to start each inductive step by a number of $\forall E$ rules and end it with a number of $\forall I$ rules.

The quantifier-reduced form of a proof is sensitive to the way we have obtained the formula: the Lexicon rule for the Lambek formula $np \backslash s$ has quantifier-reduced form $\forall x.np(x, 1) \multimap s(x, 2) \vdash_n \forall y.np(y, 1) \multimap s(y, 2)$ whereas Lexicon rule for the formula $s|np$ with principal type $(1 \rightarrow A) \rightarrow (2 \rightarrow A)$, which would normally be assigned the same axiom, has quantifier-reduced form $\forall x.np(x, 1) \multimap s(x, 2) \vdash_n np(A, 1) \multimap s(A, 2)$ (which we can obtain from the previous sequent by a single application of $\forall E$).

Proposition 5.8 *Let δ be first-order linear logic proof in long normal form which has the translation of a hybrid sequent as its conclusion. All occurrences*

of $\forall E$ and $\forall I$ of the quantifier-reduced from δ' of δ occur respectively in the following contexts.

$$\frac{\frac{\Gamma \vdash_n \forall x.[A \multimap B]}{\Gamma \vdash_n A \multimap B} \forall E \quad \Delta \vdash_p A}{\Gamma, \Delta \vdash_n B} \multimap E \qquad \frac{\frac{\Gamma, A \vdash_p B}{\Gamma \vdash_p A \multimap B} \multimap I}{\Gamma \vdash_p \forall x.[A \multimap B]} \forall I$$

Proof Given Proposition 5.5 and the fact that δ' is quantifier-reduced, all quantifiers occur in (sub)formulas of the form $\forall x.[A \multimap B]$, which corresponds to the translation of a Lambek formula. Given that δ is in long normal form, meaning that the focus shift rule is applied only to atomic formulas, so is its quantifier-reduced form δ' .

Look at an arbitrary application of the $\forall E$ rule. We show it must be part of a subproof of the form shown above on the left. After application of the $\forall E$ rule, we have the sequent $\Gamma \vdash_n A \multimap B$. The focus shift rule cannot apply, since $A \multimap B$ is not atomic and δ' is in long normal form. Therefore, by inspection of the available rules $\multimap E$ is the only rule available and we are in the case shown above.

The case for the $\forall I$ rule is similar. To obtain a formula $\Gamma \vdash_p A \multimap B$ as the premiss of the $\forall I$ rule, focus shift is excluded because we have a complex formula. The only available alternative removes the main connective as shown above on the right. \square

5.3 Hybrid proof to MILL1 proof

After this long setup, everything is in place to prove the main theorem. Thanks to the way we have defined our basic notions and translations, the proof is rather simple. We show that under the given translation, the proof rules of hybrid type-logical grammar are derived rules of MILL1. In the next section, we show the converse: that MILL1 proofs using formulas obtained from the translation correspond to proofs in hybrid type-logical grammars.

The proof is actually stronger: we show that proofs in the two systems generate the same *semantics*. This is easily verified since, as discussed in Section 4.3, the elimination (resp. introduction) rules for $/$, \backslash and $|$ correspond to the elimination (resp. introduction) rule for \multimap . The elimination rules (for $/$, \backslash , $|$ and \multimap) correspond to application and the introduction rule correspond to abstraction. The quantifier \forall is treated as semantically inert.

Lemma 5.9 *Let δ be a hybrid proof of $\Gamma \vdash A$, then there is an MILL1 proof δ^* of its translation $\|\Gamma\| \vdash \|A\|$.*

Proof We produce a unfocused proof with unification (that is, we do not distinguish between \vdash_p and \vdash_n). If desired, we can transform the proof obtained by this lemma into a focused proof by Proposition 2.4). We also produce a proof in quantifier-reduced form.

Since the lexicon/axioms rules are in beta-normal eta-long form by definition, we know from Lemma 3.22 of Kanazawa (2011) that substitution is only of type variables/atoms for type variables and never of a complex type for a type variable, so the arity of our predicate symbols in first-order linear logic is fixed.

Induction on the depth d of the proof.

If $d = 1$ we either have an axiom rule or a lexical hypothesis. In both cases, we have a sequent $x : A \vdash M^\alpha : A$, with α the principal type of M and with x of type α in the axiom case and of type $i \rightarrow i - 1$ (for the i th word) in the lexicon case. We translate the axiom by $\|A\|^{f(\alpha)} \vdash \|A\|^{f(\alpha)}$ (letting the free variables of $f(\alpha)$ become free meta-variables) and the lexical hypothesis by the axiom $\|A\|_c^{f(\alpha)} \vdash \|A\|^{f(\alpha)}$, replacing the meta-variables in $f(\alpha)$ on the left with variables and quantifying over them, making the formula on the left-hand side of the turnstile closed. Since we produce a proof in quantifier-reduced form, we do not perform the closure on the right-hand side of the turnstile (or, if you prefer, we perform the closure but immediately follow it by $\forall E$ rules for all quantifiers introduced by the closure operation).

If $d > 1$, induction hypothesis gives us proofs of the premisses of the rule and we proceed by case analysis on the last rule in the hybrid proof.

[$\backslash E$] By induction hypothesis, we have a proof δ_1 of $\Gamma \vdash \|B\|^{[C,D]}$ and a proof δ_2 of $\Delta \vdash \|B \backslash A\|^{[F,E]}$. In addition, we know by induction hypothesis that a MGU \mathfrak{s} of $\langle \Gamma; D \rangle$ and $\langle \Delta; F \rangle$ exists. Therefore, we can construct a proof of the conclusion of the $\backslash E$ rule as follows. Since G is fresh, unifying it with C is possible and produces a new substitution \mathfrak{s}' .

$$\frac{\frac{\frac{\vdots \delta_1}{\Gamma \vdash \|B\|^{[C,D]}} \quad \frac{\frac{\frac{\vdots \delta_2}{\Delta \vdash \|B \backslash A\|^{[F,E]}} \quad \Delta \vdash \forall x. \|B\|^{[x,F]} \multimap \|A\|^{[x,E]}}{\Delta \vdash \|B\|^{[G,F]} \multimap \|A\|^{[G,E]}} \quad \forall E}{\Gamma \vdash \|B\|^{[C,D]} \quad \Delta \vdash \|B\|^{[G,F]} \multimap \|A\|^{[G,E]}} \multimap E}{\mathfrak{s}'(\Gamma), \mathfrak{s}'(\Delta) \vdash \|A\|^{[\mathfrak{s}'(C), \mathfrak{s}'(E)]}} \multimap E$$

[$/E$] Symmetric.

[E] By induction hypothesis, we have a proof δ_1 of $\Gamma \vdash \|A\|B\|^{f(\beta \rightarrow \alpha)}$ and a proof δ_2 of $\Delta \vdash \|B\|^{f(\gamma)}$. We also know there is an MGU \mathfrak{s} of $\langle \Gamma; \beta \rangle$ and $\langle \Delta; \gamma \rangle$. Therefore, we can combine these two proofs using $\multimap E$ and (by Lemma 5.4) this same unification, as follows.

$$\frac{\frac{\frac{\vdots \delta_1}{\Gamma \vdash \|A\|B\|^{f(\beta \rightarrow \alpha)}} \quad \Gamma \vdash \|B\|^{f(\beta)} \multimap \|A\|^{f(\alpha)}}{\Gamma \vdash \|B\|^{f(\beta)} \multimap \|A\|^{f(\alpha)}} \quad \frac{\vdots \delta_2}{\Delta \vdash \|B\|^{f(\gamma)}} \quad \multimap E}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash \|A\|^{f(\mathfrak{s}(\alpha))}} \multimap E$$

[$\backslash I$] By induction hypothesis, we have a proof δ_1 of $\Gamma, \|B\|^{[D,C]} \vdash \|A\|^{[D,E]}$. In addition, since all principal types are balanced and the two occurrences of D occur in the translations of B and A respectively, we know there are no occurrences of D in Γ . Hence, after the $\multimap I$ rule, we satisfy the condition for the $\forall I$ rule and can extend the proof as follows.

$$\frac{\frac{\frac{\vdots \delta_1}{\Gamma, \|B\|^{[D,C]} \vdash \|A\|^{[D,E]}}{\Gamma \vdash \|B\|^{[D,C]} \multimap \|A\|^{[D,E]}} \multimap I}{\Gamma \vdash \forall x. \|B\|^{[x,C]} \multimap \|A\|^{[x,E]}} \forall I}{\Gamma \vdash \|B \backslash A\|^{[C,E]}} =_{def}$$

[$/I$] Symmetric.

[I] Induction hypothesis gives us a proof δ_1 of $\Gamma, \|B\|^{f(\beta)} \vdash \|A\|^{f(\alpha)}$, which we can extend as follows.

$$\frac{\frac{\frac{\vdots \delta_1}{\Gamma, \|B\|^{f(\beta)} \vdash \|A\|^{f(\alpha)}}{\Gamma \vdash \|B\|^{f(\beta)} \multimap \|A\|^{f(\alpha)}} \multimap I}{\Gamma \vdash \|A|B\|^{f(\beta \rightarrow \alpha)}} =_{def}$$

□

5.4 MILL1 proof to hybrid proof

Lemma 5.10 *Let δ be the MILL1 derivation of the translation of a hybrid sequent, that is, of $\|A_1\|^{[0,1]}, \dots, \|A_n\|^{[n-1,n]} \vdash \|B\|^{[0,n]}$. Then there is a hybrid proof δ^* of $x_1^{1 \rightarrow 0} : A_1, \dots, x_n^{n \rightarrow n-1} : A_n \vdash M^{n \rightarrow 0} : B$, where $M \equiv_{\beta\eta} \lambda z. (x_1 \dots (x_n z))$.*

Proof The fact that $M \equiv_{\beta\eta} \lambda z. (x_1 \dots (x_n z))$ follows immediately from the balanced occurrences of the type constants $0, \dots, n$.

Let δ be the focused MILL1 derivation of $\|\Gamma\| \vdash_p \|A\|$, or, the case being, of $\|\Gamma\| \vdash_n \|A\|$. We assume δ to be in quantifier-reduced form.

We proceed by induction on the depth d of the proof.

If $d = 1$, then there are two cases.

Lexicon If the rule was a lexical hypothesis, then it is a proof of $\|A_i\|^{[i-1,i]}$ for one of the A_i of the endsequent of the proof. By construction, we can recover the principal type α and (by Coherence) a unique β -normal η -long lambda term M of type α . Therefore, we have a hybrid proof $x_i^{i-1 \rightarrow i} : A_i \vdash M^\alpha : A_i$, with α the principal type by construction.

Axiom If the rule was an axiom then the formula A does not appear in the endsequent. We again recover the principal type α and the (eta-expanded) lambda term M from the translation function and we return the hybrid proof $x^\alpha : A \vdash M^\alpha : A$, with M the eta-expansion of x to produce a valid Axiom rule.

If $d > 1$, then we proceed by case analysis of the last rule of the proof.

[\pm] Induction hypothesis gives us the proof corresponding to the negative premiss of the rule. We return the same proof.

[$\forall E / \multimap E$] For the combination of a $\forall E / \multimap E$ rule, there are two cases to consider, depending on whether the translated formula had $/$ or \backslash as main connective. In case it was $/$, our translation unfolds as shown below. The MGU \mathfrak{s} unifies D with F (it doesn't matter here if the $\forall E$ step has been done separately: in that case x is replaced by a fresh metavariable G and the MGU unifies G with E).

$$\frac{\frac{\frac{\|\Gamma\| \vdash_n \|A/B\|^{[C,D]}}{\|\Gamma\| \vdash_n \forall x. \|B\|^{[D,x]} \multimap \|A\|^{[C,x]}} =_{def} \quad \frac{\|\Gamma\| \vdash_n \|B\|^{[D,E]} \multimap \|A\|^{[C,E]}}{\|\Gamma\| \vdash_n \|B\|^{[D,E]} \multimap \|A\|^{[C,E]}} \quad \forall E \quad \frac{\|\Delta\| \vdash_p \|B\|^{[F,E]}}{\|\Delta\| \vdash_p \|B\|^{[F,E]}}}{\mathfrak{s}(\|\Gamma\|), \mathfrak{s}(\|\Delta\|) \vdash_n \|A\|^{[\mathfrak{s}(C), \mathfrak{s}(E)]}} \multimap E$$

Lemma 5.4 guarantees that the two hybrid formulas B are indeed identical and induction hypothesis gives us a proof δ_1 of $\Gamma \vdash M^{D \rightarrow C} : A/B$ and a proof δ_2 of $\Delta \vdash N^{E \rightarrow F} : B$, which we can combine by the $/E$ rule, using the same substitution \mathfrak{s} , to produce a proof of $\Gamma, \Delta \vdash A^{\mathfrak{s}(E) \rightarrow \mathfrak{s}(C)}$ as required.

$$\frac{\frac{\vdots \delta_1}{\Gamma \vdash M^{D \rightarrow C} : A/B} \quad \frac{\vdots \delta_2}{\Delta \vdash N^{E \rightarrow F} : B}}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash (\lambda z. M(N z))^{\mathfrak{s}(E) \rightarrow \mathfrak{s}(C)} : A} /E$$

According to Lemma 4.1, we have also computed the corresponding principal type $\mathfrak{s}(E) \rightarrow \mathfrak{s}(C)$.

The case for \backslash is symmetric.

[$\multimap E$] In a quantifier reduced proof, a solitary $\multimap E$ (without preceding $\forall E$ producing the major premiss of the rule, which was treated in the previous case) originated from a formula $A|B$. We are in the following case.

$$\frac{\frac{\Gamma \vdash_n \|A|B\|^{f(\beta \rightarrow \alpha)}}{\Gamma \vdash_n \|B\|^{f(\beta)} \multimap \|A\|^{f(\alpha)}} \equiv_{def} \quad \Delta \vdash_p \|B\|^{f(\gamma)}}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash_n \|A\|^{f(\mathfrak{s}(\alpha))}} \multimap E$$

By induction hypothesis there is a proof of δ_1 of $\Gamma \vdash A|B$ (where the term M of $A|B$ has principal type $\beta \rightarrow \alpha$) and a proof δ_2 of $\Delta \vdash B$ (where the term N assigned to B has principal type γ). By Lemma 5.4, the two hybrid B formulas are identical and we can use the MGU \mathfrak{s} as the most general unifier of γ and β . We can therefore combine these proofs using the $|E$ rule and \mathfrak{s} as follows.

$$\frac{\begin{array}{c} \vdots \delta_1 \\ \Gamma \vdash M^{\beta \rightarrow \alpha} : A|B \end{array} \quad \begin{array}{c} \vdots \delta_2 \\ \Delta \vdash N^\gamma : B \end{array}}{\mathfrak{s}(\Gamma), \mathfrak{s}(\Delta) \vdash (MN)^{\mathfrak{s}(\alpha)} : A} |E$$

Producing principal type $\mathfrak{s}(\alpha)$ for this derivation.

$[\neg \circ I / \forall I]$ If it results from a translation with a pair of string formulas, we treat the combination of the $\neg \circ I$ and a $\forall I$ rule as a single step. By Proposition 5.8, we can do so without loss of generality. Such a combination can only result from the translation of a positive formula with main connective $/$ or $.$ We treat only $/I$; the case for $\backslash I$ is symmetric.

$$\frac{\frac{\Gamma, \|B\|^{[D,E]} \vdash_p \|A\|^{[C,E]}}{\Gamma \vdash_p \|B\|^{[D,E]} \neg \circ \|A\|^{[C,E]}} \neg \circ I}{\Gamma \vdash_p \forall x. \|B\|^{[D,x]} \neg \circ \|A\|^{[C,x]} \forall I} =_{def} \Gamma \vdash_p \|A/B\|^{[C,D]}$$

We can simply extend the proof δ_1 from the induction hypothesis as follows.

$$\frac{\begin{array}{c} \vdots \delta_1 \\ \Gamma, x^{E \rightarrow D} : B \vdash M^{E \rightarrow C} : A \end{array}}{\Gamma \vdash ((\lambda x.M)(\lambda z.z))^{D \rightarrow C} : A/B} /I$$

$[\neg \circ I]$ Finally, the case where the $\neg \circ I$ is not followed by a $\forall I$ corresponds to the $|I$ rule. We are in the following situation.

$$\frac{\frac{\Gamma, \|B\|^{f(\beta)} \vdash_p \|A\|^{f(\alpha)}}{\Gamma \vdash_p \|B\|^{f(\beta)} \neg \circ \|A\|^{f(\alpha)}} \neg \circ I}{\Gamma \vdash_p \|B\|A\|^{f(\beta \rightarrow \alpha)}} =_{def}$$

We can simply extend the proof δ_1 of $\Gamma, x^\beta : B \vdash M^\alpha : A$ given by the induction hypothesis as follows.

$$\frac{\begin{array}{c} \vdots \delta_1 \\ \Gamma, x^\beta : B \vdash M^\alpha : A \end{array}}{\Gamma \vdash (\lambda x.M)^{\beta \rightarrow \alpha} : A|B} |I$$

□

5.5 Main Theorem

Theorem 5.11 *Derivability of hybrid type-logical grammars and their translation into first-order linear logic coincides. Moreover, proofs in the two systems produce the same semantic lambda terms.*

Proof Immediate from Lemma 5.9 and Lemma 5.10 and the observation that $/E$, $\backslash E$ and $|E$, like $\multimap E$ to which they correspond by translation, are all translated as application on the meaning level and similarly for the different introduction rules and abstraction. □

Thanks to Theorem 5.11, we can use the well-understood proof theory of first-order linear logic for parsing/theorem proving hybrid type-logical grammars. Besides (focused) natural deduction and proof nets, discussed in Section 2, the work on sequent proof search of Lincoln & Shankar (1994), which includes a treatment of the additives, can also directly be applied. These proof systems all have their strengths and inconveniences, but, since they are all equivalent we can choose the most appropriate tool for the job. For example, focused natural deduction and proof nets simplify the work of enumerating readings for a given statement, and, as shown in Figure 4, proof nets provide an easy way to show undervability of a statement. In addition, the main theorem has the following immediate consequence.

Corollary 5.12 *Hybrid type-logical grammars are NP-complete*

Proof Hardness follows from the fact that hybrid type-logical grammars contain the Lambek calculus (the implicational fragment of the Lambek calculus was shown to be NP-complete by Savateev (2009)) — or alternatively from the fact that they contain lexicalized abstract categorial grammars (de Groote 2001). Since first-order linear logic is NP-complete, by Lemma 5.9 and the fact that the translation is linear in the size of the formulas, hybrid type-logical grammars are in NP. □

To compare hybrid type-logical grammars with lambda grammars, we first define an interesting subclass of hybrid type-logical grammars which we will show to be equivalent to lambda grammars.

Definition 5.13 *A hybrid proof is strictly separated iff for every $/I$ and $\backslash I$ rule, the subproof leading to the premiss of this introduction rule consists only of Lambek elimination rules and premisses $A \vdash A$ with A a Lambek formula (ie. a member of \mathcal{F}_1 , containing only $/$, \backslash and simple atomic formulas).*

We can enforce strict separation directly in the proof theory by splitting the \vdash symbol into \vdash_L and \vdash_λ , subscripting by \vdash_L the premisses and conclusions of the $/E$, $\backslash E$, $/I$, $\backslash I$ and axiom/hypothesis for Lambek formulas as \vdash_L , subscripting

by \vdash_λ the $|E$, $|I$ and axiom/hypothesis for formulas not in \mathcal{F}_1 and adding the inclusion rule.

$$\frac{\Gamma \vdash_L M^{E \rightarrow D} : B}{\Gamma \vdash_\lambda M^{E \rightarrow D} : B} L, \lambda$$

Not all proofs in hybrid type-logical grammars are strictly separated, as shown by the example in Section 4.2 on page 25, where the final $/I$ rule is preceded by both $|E$ and $|I$.

Lemma 5.14 *Strictly separated hybrid type-logical grammars generate the same string languages and the same string-meaning relations as lambda grammars.*

Proof (sketch) The main idea from (Buszkowski 1996), who uses a variant of the proof from (Pentus 1995, Pentus 1997), is that we can replace Lambek calculus formulas by *sets* of atomic formulas (CFG nonterminals) which behave combinatorially like AB formulas — the CFG nonterminals are essentially the names for AB formulas — in such a way that these sets generate the same lambda term semantics. Here, we do the same for all Lambek *sub*-formulas of a given hybrid type-logical grammar.

By the definition of strict separation, we know that all Lambek rules occur in subproofs where these rules are not intermingled with the lambda grammar rules. Hence, Buszkowski's construction translates these proofs of $\Gamma \vdash_L B$ into proofs of $\Gamma' \vdash_L B$ where only the $/E$ and $\backslash E$ rules are used. Then, by treating all Lambek formulas as CFG nonterminals and all instantiations of the $/E$ and $\backslash E$ rules in the grammar as CFG rules. That is, the instantiation of the $\backslash E$ rule for specific formulas A and B

$$\frac{A \quad A \backslash B}{B} \backslash E$$

becomes a non-logical rule

$$\frac{D \quad E}{C}$$

(or, if we prefer to write it as a CFG rule: $D, E \longrightarrow C$), where C is the non-terminal corresponding to formula B , D corresponds to formula A and E corresponds to the formula $A \backslash B$.

□

Lemma 5.15 *Parsing lambda grammars which are the translation of strictly separated hybrid type-logical grammars is NP-complete.*

Proof The construction of Lemma 5.14 generates, by means of the Buszkowski (1996) proof, many non-logical grammar rules. Given that such a system may not be decidable, we need to be careful. However, by the construction of (Buszkowski 1996), all non-lexicalized rules are of the form $D, E \longrightarrow F$ with D ,

E and F atomic formulas. Moreover, these atomic formulas correspond to AB formulas, such that either $D = A/B$, $E = B$ and $F = A$ or $D = B$, $F = B \setminus A$ and $F = A$ (for some Lambek formulas A and B). Therefore, we can start our proof by computing the closure of these AB subproofs in $O(n^3)$, then continue the normal lambda grammar proof, which is NP-complete. \square

It should be obvious from the proof sketch of Lemmas 5.14 and 5.15 that though strictly separated hybrid type-logical grammars generate the same string languages and string-meaning pairs as lambda grammars, hybrid type-logical grammars allow a *much* more compact specification of such grammars since we avoid a brute-force explosion of the size of the lexicon and of the number of lexical entries per word. Though I don't believe that the NP-complete problems we encounter in computational linguistics are necessarily intractable — Matsuzaki, Miyao & Tsujii (2007) show that some NP-complete problems in computational linguistics can be solved *much* more efficiently than $O(n^6)$ problems — having an exponential explosion of grammar size *followed* by an NP-complete problem is profoundly worrying for those interested in actually parsing the formalism.

It is unclear whether we can generalize the proof of Lemma 5.14 to dispense with the strict separation requirement on hybrid grammars. Allowing interleaving of the Lambek grammar and lambda grammar rules seems to require a generalization of the results of (Buszkowski 1996) to the hybrid type-logical grammar case and, unless we change the proof of the theorem considerably, this would require a type of interpolation proof for hybrid type-logical grammars, which, as we will see in Section 6.1, seems problematic for the lambda grammar part of the system. For example, looking back to the proof in Section 4.2, it is unclear how to replace the final $/I$ rule by the elimination rule for either $/$ or \setminus , besides adding $s/(np \setminus s)$ directly as an additional lexical entry for the quantifier.

Also, though it is certainly a desirable property of the hybrid system to derive $s[(s|np) \vdash s/(np \setminus s)]$ (for the given lexical lambda term), since it relates the generalized quantifier formulas to one of its standard Lambek calculus formulas, it is unclear if we actually need this type of derivation to give a natural account of the linguistic data. So the following question remains open: are there any examples of hybrid type-logical grammar analyses where there is no corresponding lambda grammar analysis? Having to resort to lexical duplication is already a problem, both from a conceptual point of view and from the point of view of parsing, but are there cases where even this doesn't suffice?

Though we will leave this question unresolved, we investigate the descriptive inadequacy of lambda grammars in Section 7.

6 Comparison

The proof nets discussed in Section 2.3 provide an insightful way to compare the different calculi discussed in this article in terms of their basic “building blocks”, seen from the point of view of first-order linear logic.

We need to be careful, since this comparison only gives *necessary* conditions to be in a certain fragment of first-order linear logic, and as such, we can use it

AB-grammar

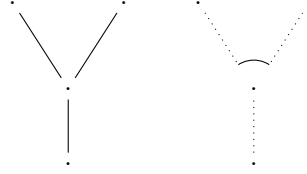


Figure 6: Lambek grammar

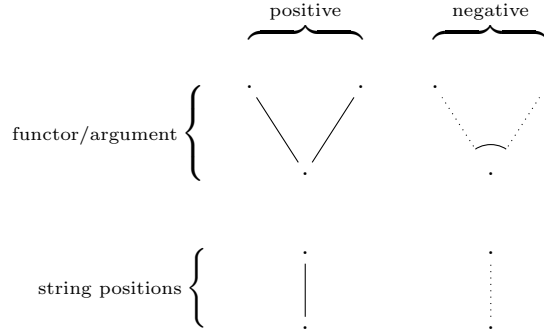


Figure 7: MILL1

only as a diagnostic for showing that possibilities are *absent* from a logic. We can directly use the different translation functions to give sufficient conditions.

The conditions on the variables in the different fragments are also absent from the visual representation. Nevertheless, we will see that this comparison is insightful.

6.1 A visual comparison of the different calculi

Figure 6 shows the Lambek calculus connectives as links for first-order linear logic proof nets. Curry’s (1961) criticism of the Lambek calculus connectives, seen from the current perspective, is that they combine subcategorization information (functor-argument structure) and string operations. Though from a modern proof-theoretical point of view (Andreoli 1992) it is perfectly valid to combine multiple positive and multiple negative rules into a single rule, separating the two gives more freedom (that is, it allows us to express more relations between the string positions and go beyond simple concatenation — the prefix and postfix of the Lambek calculus).

As shown in Figure 7, the first-order linear logic solution decomposes the

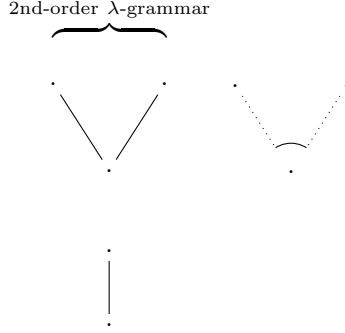


Figure 8: Lambda grammars

Lambek connectives into separate subcategorization and string position components. In a sense, this decomposition answers Curry’s critique in a very simple way.

Curry’s own solution is different and causes a loss of symmetry: as Figure 8 makes clear, the positive universal link is missing! This loss of symmetry is easy to miss in a unification-based presentation of the logic where, in addition, the quantifiers occur only as an implicit prefix of the formula. For a logician/proof theorist, this is worrying since many classical results and desirable properties of the system (restriction to atomic axioms, cut elimination, interpolation⁶) depend on this symmetry. However, it is also the cause of empirical inadequacy: positive A/B and $B \backslash A$ can no longer be represented, hence no satisfactory treatment of adverbs, coordination, gapping etc.; we will elaborate this point in detail in Section 7.

Another way to look at this is that lambda grammars require all formulas to be expressed in prenex normal form — something we exploit in the translation function. However, since we are using linear logic, not all formulas have a prenex normal form. The following are all underivable (assuming no occurrences of x in B). Refer back to Figure 4 to see why the first statement is underivable.

$$\begin{aligned}
 (\forall x.A) \multimap B &\not\vdash \exists x(A \multimap B) \\
 \exists x(A \multimap B) &\not\vdash (\forall x.A) \multimap B \\
 B \multimap \exists x.A &\not\vdash \exists x.(B \multimap A) \\
 \exists x.(B \multimap A) &\not\vdash B \multimap \exists x.A
 \end{aligned}$$

The hybrid solution to this problem is shown in Figure 9: reintroduce the positive Lambek connectives directly. There are now two ways of coding the negative Lambek connectives. The resulting system is also greater than the sum

⁶Interpolation, proved first for the Lambek calculus in (Roorda 1991) is a key component of the context-freeness proof for the Lambek calculus of Pentus (1997) and is likely to play a similar role in proofs about the generative capacity of these alternative and extended systems.

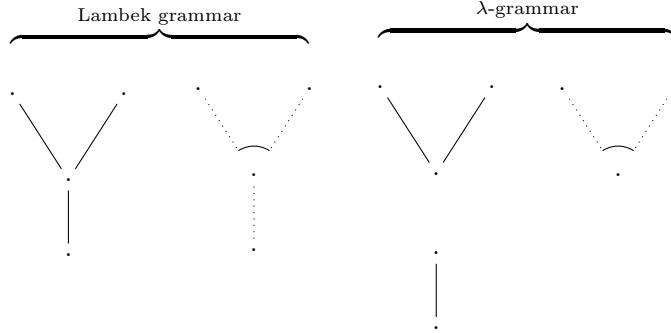


Figure 9: Hybrid grammar

of its parts, since gapping, which has a satisfactory neither in Lambek grammars nor in lambda grammars, can be elegantly treated in hybrid categorial grammar (Kubota & Levine 2012, Kubota & Levine 2013c).

Symmetry is still lost⁷, but empirically the system seems comparable to the Displacement calculus (Morrill et al. 2011): the Displacement calculus has the full symmetry absent from hybrid type-logical grammars. In spite of this, as we have seen at the end of Section 5.1, in many cases, the analyses proposed for the two formalisms basically agree, as is made especially clear by their translation into MILL1.

The differences between the two systems seems to be that hybrid type-logical grammars can, like lambda grammars, generate non-well-nested string languages and that Displacement grammars (seen from the point of view of hybrid type-logical grammars) allow the Lambek connectives to outscope the discontinuous connectives. Further analysis is necessary to decide which of these two systems has the better empirical coverage.

D grammars (Morrill et al. 2011) have a different perspective, which is shown in Figure 10. Functor argument structure and string positions are still joined, but a greater number of combinations are possible (from 0 to n quantifiers, for a small value of n determined by the grammar). Lambek grammars are now the restriction to a single quantifier for each binary connective.

D grammars enriched with bridge, left projection and right projection, shown in Figure 11, permit combinations of string position/subcategorization which are not of the same polarity. These uses are rather restricted compared to the visually similar quantifier link of first-order linear logic: essentially, they enable us to require that a pair of positions spans the empty string.

Summing up, first-order linear logic decomposes the connectives of different

⁷Neither full logical symmetry nor having the Lambek calculus as a subsystems is of course necessary to have an empirically valid formal system, as shown, for example by CCG (Steedman 2001). However it calls for further investigation as to what exactly is absent from the system and if this absence is important from a descriptive point of view. For lambda grammars, we will do this in detail in Section 7.

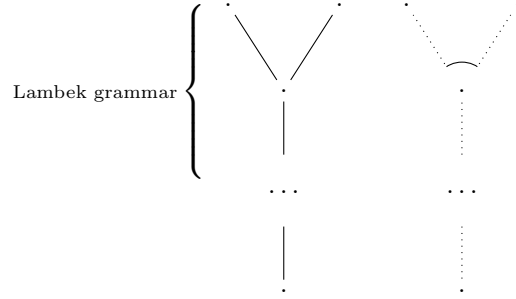


Figure 10: D grammars, binary

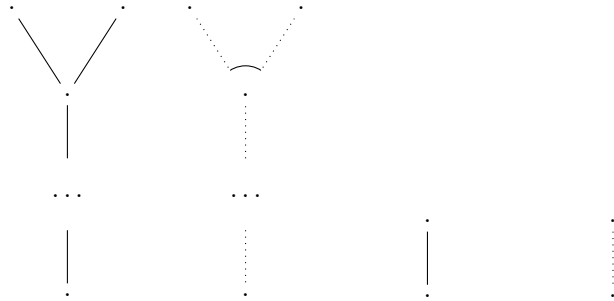


Figure 11: D grammars

grammatical frameworks — the Lambek calculus, lambda grammars, Hybrid Type-Logical Grammars and the Displacement calculus — in a natural way into its four types of links. This visual comparison both highlights the differences between this calculi and opens the way for a more detailed comparison of the descriptive limitations of one calculus compared to another.

Given that it is a decomposition of connectives, the MILL1 translation is slightly bigger in terms of the total number of connectives in the lexical entries. However, the basic operation are simple and well-understood and the first-order variables actually function as powerful constraints during proof search. Thanks to the embedding results of this paper and of (Moot 2014), we can import the large range of linguistic phenomena treated by Displacement grammars and Hybrid Type-Logical Grammar directly into MILL1.

From the point of view of first-order linear logic, the connectives of the other calculi are synthetic connectives, combined connectives of the same polarity. We can mix and match these synthetic connectives as we see fit. We can also exploit the symmetry of first-order linear logic and use lambda grammar lexical entries as arguments, restoring the symmetry of lambda grammars (and of Hybrid

Type-Logical Grammars). In addition, we can add the product \otimes and quantifier \exists to our calculus essentially for free. Moreover, as discussed in (Moot & Piazza 2001, Moot 2014) we can use the quantifiers of first-order linear logic to give an account of agreement and island constraints as well. So we can improve upon Displacement grammar analyses by adding agreement and island constraints and improve upon Hybrid Type-Logical Grammar analyses by adding symmetry, agreement and island constraints, all with the same logical primitives.

7 Descriptive Inadequacy of Lambda Grammars

As already alluded to in Section 6.1, the asymmetry of lambda grammars is the cause of descriptive inadequacy. Researcher in lambda grammars have been aware of problems with coordination at least since Muskens (2001), who briefly mentions an apparent incompatibility between lambda grammars and the categorical grammar treatment of coordination, but the problem can be traced back to (Curry 1961) where the analysis of the coordination “both ... and ...” in §5-6 is problematic. Kubota & Levine (2013*a*, 2013*c*) show how catastrophic the predictions of lambda grammars are; we will repeat their observations below while adding several additional troublesome cases. This problem has been little noted and little discussed⁸. Indeed, one can find several claims in the literature which deny there is a problem: Muskens claims elsewhere (Muskens 2003) that “Since word order is now completely encoded in the phrase structure term, there is no longer any need for a directionality of the calculus” and that “The availability of syntactic λ -terms reins in the overgeneration of the traditional undirected calculi.”. However, as we will show below, using lambda terms to limit the overgeneration of undirected calculi is only partially successful and it is exactly for this reason that a satisfactory treatment of coordination has remained elusive. Worse, the problem of overgeneration is not limited to coordination, but a problem with *any* higher-order type of the Lambek calculus. The standard higher-order lambda grammar treatments for generalized quantifiers and for non-peripheral are the only cases we know of where lambda grammars make the right predictions. But even here, the lambda grammar analysis does not generalize: generalized quantifiers can be seen as instances of Moortgat’s (1996*a*) $q(A, B, C)$ operator, and the lambda grammar treatment only works when B is atomic and therefore for quantifiers, of type $q(np, s, s)$, but not for reflexives, of type $q(np, np \setminus s, np \setminus s)$. For non-peripheral extraction, the lambda grammar analysis again presupposes the extracted element is an atomic formula and therefore the treatment does not generalize to gapping (for more on gapping see Section 7.2).

To give an idea of how widespread and serious the problems are, the following is a non-exhaustive list of problems for lambda grammars.

- (1) John deliberately hit Mary. (adverbs)

⁸At least in the lambda grammar and abstract categorical grammar literature, the problem is discussed in the context of linear grammar in (Worth 2014).

- (2) John bought a sandwich and ran to the train. (VP coordination)
- (3) John caught and ate a fish. (TV coordination)
- (4) John likes both black and gray t-shirts. (adjective coordination, after Curry, 1961)
- (5) John loves but Mary hates Noam. (right-node raising)
- (6) John bought himself a present. (reflexives)
- (7) John studies logic and Charles, phonetics. (gapping)
- (8) John left before Mary did. (ellipsis)
- (9) John ate more donuts than Mary bought bagels. (comparative sub-deletion)

These problems range from the mundane to the more involved, but the important point is that, taken together, these problems occur very frequently and that *all* cases listed above have a simple and elegant treatment in the Displacement calculus (Morrill et al. 2011), in Hybrid type-logical grammars (Kubota & Levine 2012, Kubota & Levine 2013b) and in multimodal type-logical grammars (Hendriks 1995, Kurtonina & Moortgat 1997). Sentence (1) to (5) are simply and correctly handled by Lambek grammars and Sentence (1) to (4) even by AB grammars.

Let me be precise about what I mean by descriptive inadequacy in this context, since some authors use the term with a slightly different meaning. A theory suffers from descriptive inadequacy if it fails to capture linguistic generalizations and instead has to resort to enumerating the linguistic data. In a lexicalized formalism like categorial grammars, this means we want to avoid multiplying the number of lexical entries for the words in our grammar as much as possible⁹. So in the context of the examples above, we would like Sentence (1) to use the same lexical entries as the sentence “John hit Mary”, with the lexical assignment to “deliberately” being to only addition and we would like Sentence (3) to use the same lexical entries as the sentences “John caught a fish” and “John ate a fish”, with the lexical assignment to “and” being the only difference. When I say that lambda grammars suffer from descriptive inadequacy, this does not mean that they are fundamentally unable to handle Sentences (1) to (9), since Lemma 5.14 guarantees that they can (given that the phenomena listed above all have strictly separated hybrid proofs). I mean that they cannot treat the sentences above without introducing otherwise unmotivated additional lexical entries — in fact, not without an exponential blowup of the size of the lexicon, as is clear from Lemma 5.14.

In Section 7.3 we will discuss the consequences of these problems in detail, as

⁹Maybe a more reasonable measure would prefer the *sum* of the size for all entries assigned to a word to be as small as possible, since a single entry $A_1 \oplus \dots \oplus A_n$ is not really simpler than n distinct entries A_1, \dots, A_n .

It should also be noted as the size of our grammar increases (in terms of the number of words and constructions it is able to handle), so does the size of our lexicon. So this is a relative measure rather than an absolute one.

well as some possible modifications to lambda grammars which may solve these problems, chiefly among those are extensions to hybrid type-logical grammar and to first-order linear logic.

7.1 Inhabitation machines

To show the main results, we need some additional notions of the typed lambda calculus. An inhabitation machine (see (Barendregt, Dekkers & Statman 2013)) is a type of grammar which, given a type, enumerates all possible terms of this type. Their use for categorial grammars has been pioneered by van Benthem (1995).

From page 33 of (Barendregt et al. 2013), the following two-level grammar (defined on type-context pairs) enumerates all closed inhabitants in beta-normal eta-long form of a given type.

Γ is a context, Γ, x^α denotes $\Gamma \cup \{x^\alpha\}$ (where x is distinct from the terms in Γ , so the result is again a valid context), A is an atomic type, α, β are arbitrary types and $\vec{\alpha} \rightarrow \beta$ is short for $\alpha_1 \rightarrow \dots \rightarrow \alpha_n \rightarrow \beta$.

$$\begin{aligned} L(A; \Gamma) &\implies xL(\alpha_1; \Gamma) \dots L(\alpha_n; \Gamma) && \text{if } x : \vec{\alpha} \rightarrow A \in \Gamma \\ L(\alpha \rightarrow \beta; \Gamma) &\implies \lambda x^\alpha. L(\beta; \Gamma, x^\alpha) \end{aligned}$$

The lambda grammar case is considerably more restricted: the lexical lambda terms must be linear and contain, for a given word w with corresponding variable w_s , a single occurrence of w_s . That is, we start with $\Gamma = \{w_s^{\sigma \rightarrow \sigma}\}$ and for the application rule, we partition $\Gamma - \{x^{\vec{\alpha} \rightarrow A}\}$ into jointly exhaustive, pairwise disjoint subsets and divide these over the different subterms. In addition, we want our lexical term to produce the correct word order and to be compatible with the syntactic lambda grammar derivation.

7.2 Problems for lambda grammars

In this next section, we will show several problematic cases for lambda grammars, using inhabitation machines to exhaust all possible solutions and find all of them inadequate.

Adverbs

As a first problem for lambda grammars, the Lambek calculus formula of an adverb such as “deliberately”, as it occurs in a sentence like “Eduardo deliberately fell”, is $(np \backslash s) / (np \backslash s)$ — it modifies a verb having taken all arguments except its subject and this verb phrase is on the immediate right of the adverb. If we translate this formula to a first-order formula and move (where possible) the quantifiers to the prefix and eliminate them, we obtain the formula $(\forall c. np(c, 2) \multimap s(c, D)) \multimap np(E, 1) \multimap s(E, D)$ but we cannot use the principal type $((2 \rightarrow c) \rightarrow D \rightarrow c) \rightarrow (1 \rightarrow E) \rightarrow D \rightarrow E$ (with c a fresh type constant) since it is uninhabited.

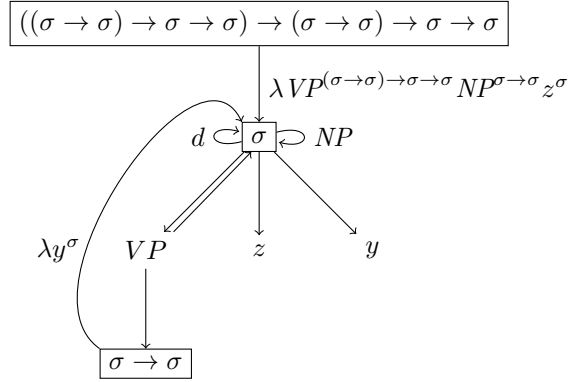


Figure 12: Inhabitation machine for an adverb type.

The lambda grammar syntactic type $(s|np)|(s|np)$ translates to the prosodic type $((\sigma \rightarrow \sigma) \rightarrow \sigma \rightarrow \sigma) \rightarrow (\sigma \rightarrow \sigma) \rightarrow \sigma \rightarrow \sigma$ and produces the inhabitation machine shown in Figure 12. We use the variable d (of type $\sigma \rightarrow \sigma$) to stand for the occurrence of the string “deliberately”. We can see that the VP node in the figure requires first an argument of type $\sigma \rightarrow \sigma$ (the downward arrow) then an argument of type σ (the upward arrow) to produce a term of type σ . Valid linear paths through the machine must pass each term label exactly once, and must pass the λy -label (on the curved arrow upwards to σ) before the y variable.

Figure 13 spits the σ node in two, making the scope of the y variable clearer.

The word order of the sentence constrains the paths we can take. We must take an NP arc before we take a d arc, since “deliberately” occurs after the subject noun phrase. So from the top σ node, we can only take three possible paths, as shown below. For comparison, the uninhabited type corresponding most closely to the first-order formula is shown as item 4. We can see that the three other types are obtained by replacing the c constant by a C variable and exchanging one of the occurrences of C with another atomic type in such a way that the resulting type is inhabited.

- (1) $\lambda VP \lambda NP \lambda z. NP (d ((VP \lambda y. y) z)) :$
 $((C \rightarrow C) \rightarrow D \rightarrow 2) \rightarrow (1 \rightarrow E) \rightarrow D \rightarrow E$
- (2) $\lambda VP \lambda NP \lambda z. NP ((VP \lambda y. d y) z) :$
 $((2 \rightarrow 1) \rightarrow D \rightarrow C) \rightarrow (C \rightarrow E) \rightarrow D \rightarrow E$
- (3) $\lambda VP \lambda NP \lambda z. ((VP \lambda y. NP (d y)) z) :$
 $((2 \rightarrow C) \rightarrow D \rightarrow E) \rightarrow (1 \rightarrow C) \rightarrow D \rightarrow E$
- (4) *Uninhabited:*
 $((2 \rightarrow c) \rightarrow D \rightarrow c) \rightarrow (1 \rightarrow E) \rightarrow D \rightarrow E$

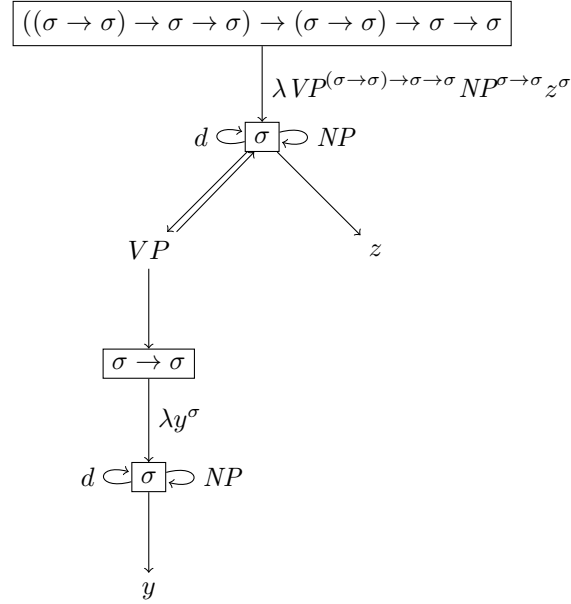


Figure 13: Simplified inhabitation machine for an adverb type.

We investigate the three possibilities in turn.

Lambda term 1 comes closest to the first-order linear logic formula, but it is a lambda term modeled after those used for extraction and, as such, it takes a sentence missing a noun phrase *anywhere* as its argument, instead of a verb phrase. Therefore, it incorrectly predicts that the three following sentences are all grammatical.

- (10) John deliberately Mary hit.
- (11) John deliberately Mary insinuates likes Susan.
- (12) John deliberately Mary hit the sister of.

Predicting that sentence (10) means “It was deliberate on the part of John that Mary hit him”, with sentence (11) meaning approximately “John made Mary insinuate that he likes Susan” and sentence (12) meaning something like “Mary hit the sister of John and this was deliberate on the part of John”. It seems very difficult to block this example without also blocking the noun “boy which Mary likes the sister of” (not super-natural, but we want to allow these kinds of extractions which are essentially indistinguishable from the current formula).

Lambda term 2 shifts from the extraction-like lambda term and its corresponding overgeneration to a lambda term similar to those used for in situ

binding/quantifying in¹⁰, where we require a sentence missing a noun phrase at the position of “deliberately” as argument. Though this analysis again allows us to derive the correct word order, it also makes the dubious claim that there is an *np* constituent at the position of the adverb. In addition, it overgenerates as follows.

- (13) Mary John hit deliberately.
- (14) Mary the friend of deliberately left.
- (15) Mary John gave the friend of deliberately a book.

Though it is possible to argue that sentence (13) is a sort of topicalization (with stress on *Mary*), it is problematic that this topicalization is triggered by the adverb, since topicalization is independent of the presence or absence of adverbs. Moreover, we generate the semantics “It was deliberate on the part of Mary that John hit her” for sentence (13). We generate the semantics “It was deliberate on the part of Mary that her friend left” for sentence (14) and similarly “Mary incited John to give her friend a book” for sentence (15).

Finally, lambda term 3 selects for a sentence missing a noun phrase with the only condition that this noun phrase occurs directly before the adverb. Here, we make the odd claim that the noun phrase and the adverb *together* span the position of an *np*: that is, it claims that an adverb is a post-modifier of an *np*. In addition, it is again an in situ binding/quantifying in analysis, but this time with the complex string “*np* deliberately” (where lambda term 2 used an in situ binding analysis of just the word “deliberately”).

- (16) John hit Mary deliberately.
- (17) The friend of Mary deliberately left.
- (18) The friend of Mary deliberately who lives in Paris left.

Though sentences (16) and (17) are syntactically correct, the problem is that we generate the semantics “It was deliberate on the part of Mary that John hit her” for sentence (16) and a reading “It was deliberate on the part of Mary that her friend left” for sentence (17) and (18).

In sum, we cannot capture the essence of the Lambek calculus formula $(np \backslash s) / (np \backslash s)$ in lambda grammars. Other adverb formulas — $(np \backslash s) \backslash (np \backslash s)$ (an adverb occurring after the verb phrase) and $(n/n) / (n/n)$ (for adverbs such as “very”), etc. — suffer from the same problem. The best approximations that we can obtain all suffer from overgeneration because non-commutativity is insufficiently enforced.

There is, of course, a solution which replaces the complex $np \backslash s$ argument by a new atomic formula, say *vp* and then, for all lexical items of the form $((np \backslash s) / A_n) \dots / A_1$, adds an additional formula $(vp / A_n) \dots / A_1$. This would essentially double the number of lexical formulas for verbs, adverbs and prepo-

¹⁰As we have seen, a generalized quantifier like “everyone” is assigned the lambda term $\lambda P.P(e)$ with *e* being the string constant corresponding to the word “everyone”.

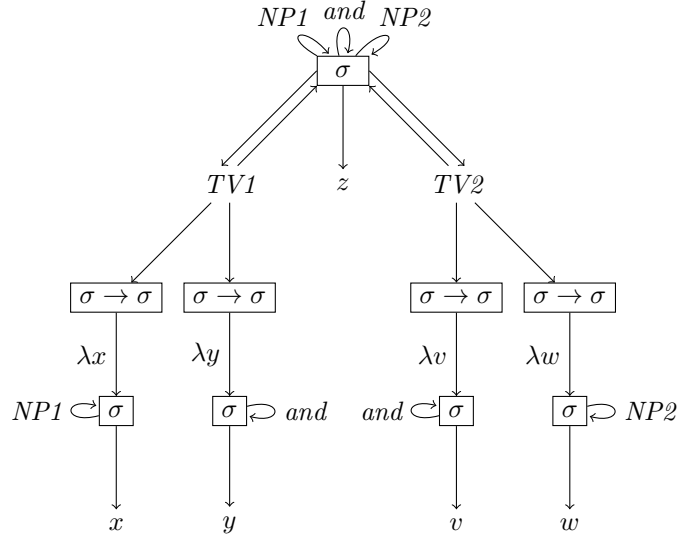


Figure 14: Simplified inhabitation machine for transitive verb conjunction

sitions — syntactic categories which already have a high number of lexical formulas — for just a single type of problematic example... More such examples will follow.

We will discuss this potential solution in a bit more detail in Section 7.3, but it should already be clear that this is not a particularly attractive option, since it is a prototypical example of descriptive inadequacy, the reasons for doubling the lexicon are purely theory-internal: no other categorial grammar, not even AB grammars, have this type of overgeneration for the simple cases we’ve shown.

Coordination

As noted by Kubota & Levine (2013a, 2013c), we can play a similar game for “John caught and ate a fish”, which looks as shown in Figure 14; for the sake of space, we do not show the prefix $\lambda TV2.\lambda TV1.\lambda NP2.\lambda NP1.\lambda z$, where $TV2$ is the transitive verb to the right of “and” (“ate” in the current example), $TV1$ is the transitive verb to the left of it (“caught”), $NP1$ is the subject, $NP2$ is the object and z is the end of the complete string.

Remark that “and” takes all constituents as argument: the two transitive verbs, the subject noun phrase and the object noun phrase, so it would seem that we *should* be able to generate the right string.

As before, we have split the σ and $\sigma \rightarrow \sigma$ nodes for readability; the actual graph merges all σ and all $\sigma \rightarrow \sigma$ nodes. The implausible analyses with $TV1$ and subject of $TV2$ and with $TV2$ as object of $TV1$ are not shown in the figure, but they fail for the same reasons discussed below.

The graph of Figure 14 shows that the *TV1* node takes first its subje^t (down and to the left of it), then its object (directly below) and finally an argument of type σ (the upward arrow back to σ) and similarly for *TV2*. The *TV1* node (optionally) takes *NP1* as its subje^t and *TV2* (optionally) takes *NP2* as its object.

Two combinations are fairly limited: the second argument of *TV1* is either *NP1* or the empty string and the first argument of *TV2* is either *NP2* or the empty string. However, if the lexical entry contains the subterm $((TV1 M) NP1)$ (for some *M* at the place of the object), then we are essentially using a quantifying-in analysis for the subject: $(TV1 M)$ is a sentence missing a noun phrase *anywhere* and applying this term to an argument puts this argument back at the place of the missing noun phrase. Consequently, it would allow the derivation of “caught John and ate a fish”. Similar overgeneration occurs for “ate” and “a fish” if we use the quantifying-in combination $(TV2 NP2)$ for the object.

If we want to avoid both types of overgeneration (subject quantifying-in and object quantifying in), the only remaining analysis consists of choosing $\lambda x.x$, $\lambda y.y$, $\lambda v.v$ and $\lambda w.w$ as arguments for the two transitive verbs.¹¹ This solution is shown in full below.

$$\lambda TV2.\lambda TV1.\lambda NP2.\lambda NP1.\lambda z. \\ NP1((TV1 \lambda x.x \lambda y.y)(and((TV2 \lambda v.v \lambda w.w)(NP2 z))))$$

As we can see from the proof in Figure 15, this lexical type allows us to derive “John caught and ate a fish” with the correct semantics. The proof has been slightly simplified by using distinct variables for the words instead of complex lambda terms (ie. we have not done lexical substitution). This has the advantage that we can use the resulting lambda term for computing the semantics as well, for which we use the following (standard) semantic substitutions¹². We can obtain the prosodic lambda terms from the principal types and the string positions (eg. $(2 \rightarrow 1) \vdash (B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A$ for “caught”, which is the standard transitive verb principal type we have seen before). The semantic terms below are all standard.

$$\begin{aligned} and &= \lambda TV1 \lambda TV2 \lambda y \lambda x. ((TV1 y) x) \wedge ((TV2 y) x) \\ j &= john' \\ f &= a_fish' \\ c &= caught' \\ a &= ate' \end{aligned}$$

¹¹This solution still overgenerates because it equates transitive verb with “sentence missing two noun phrases” and therefore incorrectly predicts that “John likes \square_{np} ’s friend from \square_{np} ” can felicitously fill this role as follows.

(i) Mary went to and John likes ’s friend from Paris.

Meaning “Mary went to Paris and John likes Mary’s friend from there”.

¹²To keep this example simple, we have treated “a fish” as an individual constant instead of a quantified noun phrase, since quantification is irrelevant for this example.

Figure 15: Proof of “John caught and ate a fish” (simplified)

$$\begin{array}{c}
 \frac{\frac{John}{np^{1 \rightarrow 0}}}{\frac{(((and\ a)\ c)\ f)\ j)^{5 \rightarrow 0}}{s}} \\
 \frac{\frac{\frac{caught}{c^{(B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A}}}{(s|np)|np}}{\frac{((and\ a)\ c)^{(L \rightarrow 4) \rightarrow (1 \rightarrow K) \rightarrow L \rightarrow K}}{(s|np)|np}} \quad \frac{\frac{\frac{and}{and^{((G \rightarrow G) \rightarrow (H \rightarrow H) \rightarrow J \rightarrow 3) \rightarrow ((E \rightarrow E) \rightarrow (F \rightarrow F) \rightarrow 2 \rightarrow I) \rightarrow (L \rightarrow J) \rightarrow (I \rightarrow K) \rightarrow L \rightarrow K}}}{(((s|np)|np) | ((s|np)|np)) | ((s|np)|np)}}{\frac{((s|np)|np) | ((s|np)|np)}{(s|np)|np}} \quad \frac{\frac{ate}{a^{(D \rightarrow 4) \rightarrow (3 \rightarrow C) \rightarrow D \rightarrow C}}}{(s|np)|np} \\
 \frac{((and\ a)\ c)^{(L \rightarrow 4) \rightarrow (1 \rightarrow K) \rightarrow L \rightarrow K}}{(s|np)|np} \quad \frac{a\ fish}{f^{5 \rightarrow 4} np} \\
 \frac{(((and\ a)\ c)\ f)^{(1 \rightarrow K) \rightarrow 5 \rightarrow K}}{s|np}
 \end{array}$$

Unfortunately, this analysis of “and” also make the (rather catastrophic) prediction that “John caught and ate a fish” has a second reading which can be paraphrased as “John caught a fish and a fish ate John”. This reading is easy to miss when we look only at eta-short proofs, since the key point of this second derivation involves switching the two arguments of the transitive verb, as shown in Figure 16.¹³ The crux of this second proof is that swapping the two arguments of “ate” is a purely local operation which has no visible effects on the word order: the only difference between the proof in Figure 15 and the proof in Figure 16 is in the subproof with undischarged hypothesis “ate” (with term a resp. $\lambda x.\lambda y.(a\ y)\ x$).

As shown in the figure, the second proof computes the following “deep structure”.

$$(((and(\mathbf{C}a))\ c)\ f)\ j = (((and\ \lambda x.\lambda y.((a\ y)\ x)\ c)\ f)\ j$$

In a similar way, we can obtain a third and a fourth reading, corresponding the string “John caught and ate a fish” but to the meanings “A fish caught John and John ate a fish” and “A fish caught and ate John” respectively, as follows.

$$\begin{aligned} (((and\ a)\ (\mathbf{C}c))\ f)\ j &= (((and\ a)\ \lambda v.\lambda w.(c\ w)\ v)\ f)\ j \\ (((and(\mathbf{C}a))\ (\mathbf{C}c))\ f)\ j &= (((and\ \lambda x.\lambda y.((a\ y)\ x)\ \lambda v.\lambda w.(c\ w)\ v)\ f)\ j \end{aligned}$$

The problem is that though we would *want* the two $(s|np)|np$ arguments of “and” to be transitive verbs, they *mean* “a sentence missing two np arguments anywhere”, which is what causes the problems with commutativity.

We can again remedy this by adding new lexical entries, for example choosing tv for the two transitive verbs and $(tv\ ((np\ s)/np))/tv$ for the conjunction, but this would mean adding several other lexical entries to analyse sentences like “John has understood and will probably implement Dijkstra’s algorithm”, which are handled by the Lambek calculus analysis — since “has understood” and “will probably implement” can both be analysed as $(np\ s)/np$ — but not by the new atomic tv analysis. So adding lexical entries is not only inelegant and an admittance of descriptive inadequacy, but such additions can cascade throughout the grammar.

I would seem that another simple potential solution would be to add case to lambda grammars. While adding case to first-order linear logic is something we can do essentially for free using extra arguments, adding case to lambda grammars at least complicates either the grammars or the types. In addition, though case would exclude the subject-object swaps we have seen in this section, it is easy to see this would not be a real solution, because the sentences in (19) below are all sentences missing a subject/nominative np , those in (20) sentences missing an object/accusative np and those in (21) sentences missing both a subject and an object (for clarity, the missing subjects and objects have been shown as $\boxed{}$ _s and $\boxed{}$ _o respectively). So while adding case excludes some

¹³The term $\lambda f.\lambda x.\lambda y.((f\ y)\ x)$ which switches subject and object is of course the \mathbf{C} combinator we have already seen in Section 3.3. It commutes the two arguments of a function f , and the proof shown in Figure 16 has a subproof which computes $\mathbf{C}a \equiv \lambda x.\lambda y.(a\ y)\ x$.

[illegible]

bad derivations, we would still predict sentences like “*Sue likes Mary and John saw the man whom likes” is grammatical (with meaning “Sue likes Mary and John saw the man whom Sue likes.”), that “ *John saw the friend of who lives in Paris and Ted likes Sue” is grammatical and means “John saw the friend of Sue who lives in Paris and Ted likes Sue” and that “*Sue John believes avoids but Ted saw whom kissed Peter” is grammatical and means “John believes Sue avoids Peter but Ted saw Peter whom Sue kissed”.

- (19) a. \Box_s likes Mary.
b. John believes \Box_s left.
c. John saw the man whom \Box_s likes.
- (20) a. John likes \Box_o .
b. John saw the friend of \Box_o who lives in Paris.
c. Ted gave \Box_o flowers.
- (21) a. \Box_s gave \Box_o flowers.
b. John believes \Box_s avoids \Box_o .
c. John saw \Box_o whom \Box_s kissed.

While it would certainly be possible to appeal to island constraints or other independently motivated mechanisms to exclude coordination of the phrases listed above, it seems that use case for this purpose is inherently on the wrong track: it uses lambda terms to encode word order for negative implications and a cascade of stop-gap solutions to constrain word order for positive implications. As the examples above make clear, for coordination, we don’t coordinate (partial) constituents which have the same case marking, but rather those which have the same *structure* and Lambek calculus formulas, such as $(np \backslash s)/np$ for transitive verb conjunction, are a good proxy for this notion of the same structure.

Though we have given an in-depth analysis only of transitive verb conjunction, other conjunctions of complex types (adjectives, intransitive verbs, etc.) suffer similar problems. In all cases, the lambda grammar analysis is between a rock and a hard place, suffering either from overgeneration (as the adverb case) or from bizarre readings (as in the transitive verb conjunction case).

Gapping

As a last problem case, the standard (multimodal) categorial grammar analysis of gapping (Hendriks 1995), of which we have seen the hybrid version in Section 5.1, does not fare any better when we try to translate it into lambda grammar. The analysis of a sentence like

- (22) John studies logic and Charles phonetics.

would assign “and” the formula.

$$((s|((s|np)|np))|(s|((s|np)|np)))|(s|((s|np)|np))$$

The idea behind the analysis of Hendriks (1995) is that “and” takes first two sentences missing a transitive verb as its arguments, then a transitive verb

to produce a sentence by placing the transitive verb back to its normal place in the first argument (which is the sentences to its left missing a transitive verb) and using the empty string instead of the transitive verb in the second sentence. In short, it uses a quantifying-in analysis for the transitive verb in the sentence to the left and an extraction analysis for the “missing” transitive verb in the sentence to the right. What is nice about this analysis, is that we use a normal coordination formula for “and”, an instance of the schema $(X|X)|X$ with (in this case) $X = (s|np)|np$. Since the analysis of Hendriks (1995) uses a combination of quantifying in and extraction, it is tempting to think that the lambda grammar analysis is unproblematic. However, the key point of the analysis is that we need both extraction and in situ binding for a complex formula, a transitive verb, though unlike for the coordination case it occurs in a negative position in the gapping coordination type. Let’s investigate the possibilities.

The formula for transitive verb gapping produces the (simplified and reduced) inhabitation machine shown in Figure 17.

As before, the prefix $\lambda STV2.\lambda STV1.\lambda TV.\lambda z.$ has been removed from the figure; $STV1$ denotes the sentence missing a transitive verb to the left of “and” (“John logic” in our case) and $STV2$ denotes the sentence missing a transitive verb to the right of “and” (“Charles phonetics” in our case), TV the transitive verb (here: “studies”) and z the end of the string.

We can obtain the full combinatorics by identifying all nodes with the same type. The current reduced graph emphasizes the reasonable lambda terms: for example, TV can only be an argument of $STV1$, corresponding to the the quantifying-in analysis producing the desired word order “John studies logic”, similarly, the first argument of the transitive verb has been restricted to the subject and the second argument to the object. In fact, getting the word order and semantics right leaves a unique lambda term — this is just the term from (Bourreau 2013), where the subterm $(STV1 TV)$ has been eta-expanded.¹⁴

$$\lambda STV2.\lambda STV1.\lambda TV.\lambda z.(((STV1 \lambda O1 \lambda S1 \lambda x.(((TV \lambda w.O1 w) \lambda v.S1 v)) x) \\ (and (STV2 \lambda O2 \lambda S2 \lambda y.S2(O2 y)) z)))$$

The principal type of this term is.

$$(((L \rightarrow K) \rightarrow (K \rightarrow M) \rightarrow L \rightarrow M) \rightarrow J \rightarrow 4) \rightarrow \\ (((D \rightarrow C) \rightarrow (F \rightarrow E) \rightarrow H \rightarrow G) \rightarrow 3 \rightarrow I) \rightarrow \\ ((D \rightarrow C) \rightarrow (F \rightarrow E) \rightarrow H \rightarrow G) \rightarrow J \rightarrow I$$

Given this lambda term and principal type, we can derive the correct word order and semantics as shown in Figure 18. We have again abbreviated the

¹⁴The eta-short term looks as follows.

$$\lambda STV2.\lambda STV1.\lambda TV.\lambda z.((STV1 TV) \\ (and (STV2 \lambda O2 \lambda S2 \lambda y.S2(O2 y)) z)))$$

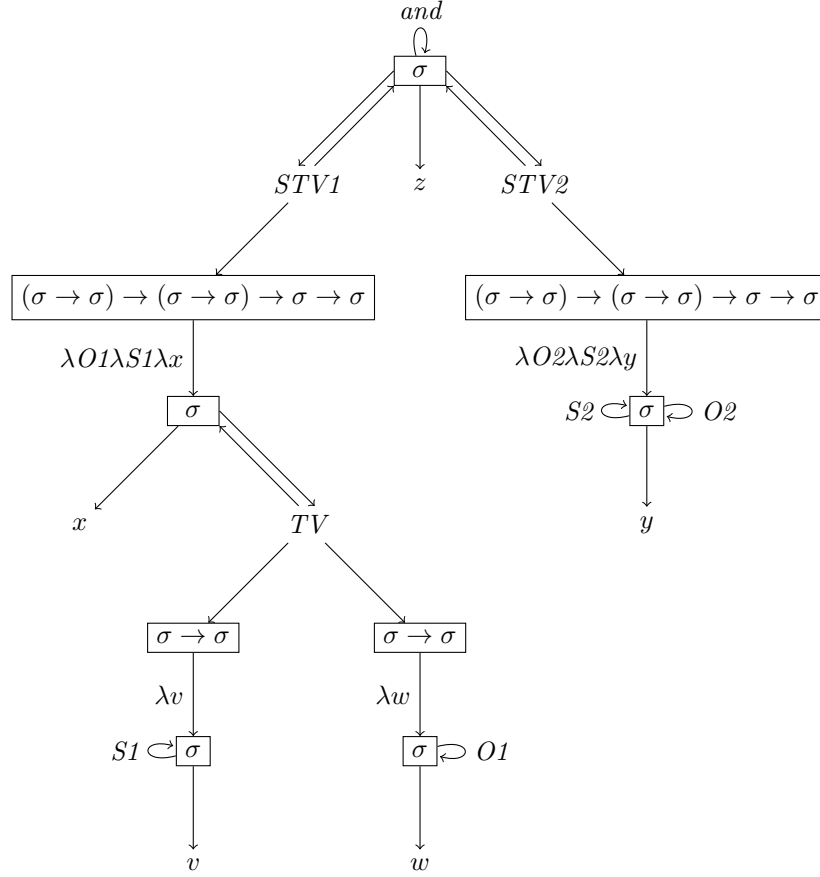


Figure 17: Simplified inhabitation machine for gapping.

proof, using the following abbreviations for readability.

$$\begin{aligned}
 X &= s|((s|np)|np) \\
 \alpha &= ((L \rightarrow K) \rightarrow (K \rightarrow M) \rightarrow L \rightarrow M) \rightarrow J \rightarrow 4 \\
 \beta &= (D \rightarrow C) \rightarrow (F \rightarrow E) \rightarrow H \rightarrow G
 \end{aligned}$$

We have also performed the substitutions necessary for the $\rightarrow E$ rules directly on the hypotheses of the proof. We obtain the semantics by substituting the

following terms for the constants in the lambda term computed for this proof.

$$\begin{aligned}
and &= \lambda STV1 \lambda STV2 \lambda TV. (STV1 TV) \wedge (STV2 TV) \\
j &= john' \\
l &= logic' \\
c &= charles' \\
p &= phonetics' \\
s &= studies'
\end{aligned}$$

But again, there is an alternative proof, shown in Figure 19. This proof swaps both the arguments of P and the two abstractions of s (“studies”). The net result is that the left conjunct stays as before, syntactically and semantically, since the two swaps cancel out against each other. Now the right conjunct has its arguments swapped in the semantics only, giving the absurd reading “John studies logic and phonetics studies Charles”.

Since this article is already rather long, we cannot treat the other problem cases mentioned at the start of Section 7. However, the examples which have been treated in detail serve as a blueprint to constructing similar problems for the additional listed problem cases. In all cases, the fundamental asymmetry of lambda grammars means we have insufficient tools at our disposal to constrain the word order for positive implications and that the best possible approximations are inadequate both syntactically and semantically.

7.3 Solutions for lambda grammars

Given the descriptive challenges for lambda grammars, it seems natural to ask how lambda grammars could evolve to rise to these challenges.

1. *Stasis*. Keeping the formalism and the analyses as they are is a possible, if not a very attractive solution, since it would require us to significantly tone down the ambitions of the syntax-semantics interface of the formalism, thereby losing one of the attractive aspects of categorial grammars.

We can also choose to embrace descriptive inadequacy and use the result from (Buszkowski 1996), which, as discussed in Section 5.5, translates Lambek grammars into AB grammars while preserving the semantics, though at the price of an explosion in lexicon size (as Pentus’, 1995 original proof) — to obtain at least the most of the coverage of hybrid type-logical grammars directly within lambda grammars (though this presupposes strict separation, as required for the application of Lemma 5.14). This would save lambda grammars empirically, incorporating the syntax-semantics interface of the hybrid system, but we would then have a combinatorial explosion followed by an NP-complete problem (according to Lemma 5.15). Given that the original Pentus proof, with $O(|G|n^3)$ complexity for some colossal $|G|$, never resulted in fast, practical parsers for

Figure 18: Proof of “John studies logic and Charles phonetics”.

$ \begin{array}{c} \text{John} \\ \hline \begin{array}{c} P^{(3 \rightarrow 2) \rightarrow (1 \rightarrow 0) \rightarrow B' \rightarrow A'} \\ (s np) np \\ \hline j^{1 \rightarrow 0} \\ np \end{array} \end{array} $	$ \begin{array}{c} \text{Charles} \\ \hline \begin{array}{c} C^{5 \rightarrow 4} \\ np \end{array} \end{array} $	$ \begin{array}{c} \text{phonetics} \\ \hline \begin{array}{c} Q^{(6 \rightarrow 5) \rightarrow (5 \rightarrow 4) \rightarrow D' \rightarrow C'} \\ (s np) np \\ \hline p^{6 \rightarrow 5} \\ np \end{array} \end{array} $
$ \begin{array}{c} \hline \begin{array}{c} (P l) j^{(1 \rightarrow 0) \rightarrow B' \rightarrow A'} \\ s np \\ \hline ((P l) j)^{B' \rightarrow A'} \\ s \\ \hline (\lambda P. (P l) j)^{((3 \rightarrow 2) \rightarrow (1 \rightarrow 0) \rightarrow B' \rightarrow A') \rightarrow B' \rightarrow A'} \\ s ((s np) np) \\ \hline ((and (\lambda Q. (Q p) c)) (\lambda P. (P l) j))^{((3 \rightarrow 2) \rightarrow (1 \rightarrow 0) \rightarrow 3 \rightarrow I) \rightarrow 6 \rightarrow I} \\ s ((s np) np) \\ \hline (((and (\lambda Q. (Q p) c)) (\lambda P. (P l) j)) (\lambda y x. (s y) x))^{6 \rightarrow 0} \\ s \end{array} \end{array} $	$ \begin{array}{c} \hline \begin{array}{c} and \\ and^{\alpha \rightarrow (\beta \rightarrow 3 \rightarrow I) \rightarrow \beta \rightarrow J \rightarrow I} \\ (X X) X \\ \hline (and (\lambda Q. (Q p) c))^{(\beta \rightarrow 3 \rightarrow I) \rightarrow \beta \rightarrow 6 \rightarrow I} \\ X X \\ \hline (((and (\lambda Q. (Q p) c)) (\lambda P. (P l) j)) (\lambda y x. (s y) x))^{6 \rightarrow 0} \\ s \end{array} \end{array} $	$ \begin{array}{c} \hline \begin{array}{c} studies \\ s^{(B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A} \\ (s np) np \\ \hline (s y)^{(1 \rightarrow A) \rightarrow B \rightarrow A} \\ s np \\ \hline ((s y) x)^{B \rightarrow A} \\ s \\ \hline (\lambda x. (s y) x)^{(1 \rightarrow A) \rightarrow B \rightarrow A} \\ s np \\ \hline (\lambda y x. (s y) x)^{(B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A} \\ (s np) np \\ \hline \end{array} \end{array} $

Figure 19: Proof of “John studies logic and Charles phonetics” with semantics “John studies logic and phonetics studies Charles”.

$$\begin{array}{c}
\frac{\frac{\frac{John}{j^{1 \rightarrow 0}} \quad \frac{P^{(1 \rightarrow 0) \rightarrow (3 \rightarrow 2) \rightarrow B' \rightarrow A'}}{np} \quad (s|np)|np}{(Pj)^{(3 \rightarrow 2) \rightarrow B' \rightarrow A'}} \quad \frac{logic}{l^{3 \rightarrow 2} \quad np}}{\frac{(Pj)l^{B' \rightarrow A'}}{s}} \\
\frac{(\lambda P.(Pj)l)^{((1 \rightarrow 0) \rightarrow (3 \rightarrow 2) \rightarrow B' \rightarrow A') \rightarrow B' \rightarrow A'}}{s|((s|np)|np)} \\
\hline
((and(\lambda Q.(Qp)c))(\lambda P.(Pj)l))^{((1 \rightarrow 0) \rightarrow (3 \rightarrow 2) \rightarrow 3 \rightarrow I) \rightarrow 6 \rightarrow I} \\
s|((s|np)|np)
\end{array}
\quad
\begin{array}{c}
\frac{\frac{Charles}{c^{5 \rightarrow 4}} \quad \frac{\frac{Q^{(6 \rightarrow 5) \rightarrow (5 \rightarrow 4) \rightarrow D' \rightarrow C'}}{(s|np)|np} \quad \frac{phonetics}{p^{6 \rightarrow 5} \quad np}}{(Qp)^{(5 \rightarrow 4) \rightarrow D' \rightarrow C'}} \quad \frac{s|np}{s|np}}{((Qp)c)^{D' \rightarrow C'}} \\
\frac{and \quad \frac{and^{\alpha \rightarrow (\beta \rightarrow 3 \rightarrow I) \rightarrow \beta \rightarrow J \rightarrow I}}{(X|X)|X} \quad \frac{(\lambda Q.(Qp)c)^{((6 \rightarrow 5) \rightarrow (5 \rightarrow 4) \rightarrow D' \rightarrow C') \rightarrow D' \rightarrow C'}}{s|((s|np)|np)}}{X|X} \\
\hline
(((and(\lambda Q.(Qp)c))(\lambda P.(Pj)l))(\lambda xy.(s y)x))^{6 \rightarrow 0} \\
s
\end{array}
\quad
\begin{array}{c}
\frac{\frac{studies}{s^{(B \rightarrow 2) \rightarrow (1 \rightarrow A) \rightarrow B \rightarrow A}} \quad \frac{y^{B \rightarrow 2} \quad np}{(s|np)|np}}{(sy)^{(1 \rightarrow A) \rightarrow B \rightarrow A}} \\
\frac{x^{1 \rightarrow A} \quad np}{np} \quad \frac{(sy)x^{B \rightarrow A}}{s} \\
\frac{(\lambda y.(sy)x)^{(B \rightarrow 2) \rightarrow B \rightarrow A}}{s|np} \\
\hline
(\lambda xy.(sy)x)^{(1 \rightarrow A) \rightarrow (B \rightarrow 2) \rightarrow B \rightarrow A} \\
(s|np)|np
\end{array}$$

the Lambek calculus because of the grammar size constant, having a similar constant for an NP-complete problem does not bode well for parsing the resulting grammar.

So it seems we have two unappealing options here: give up — or significantly reduce the ambition of — the syntax-semantics interface or give up actually parsing lambda grammars.

2. Change the *terms* and/or their *interpretation*. The lambda grammars discussed above produce strings. Several authors have looked at lambda grammars which generate different types of structures, such as trees (Muskens 2001, de Groote 2002). When we generate trees, we can add a separate yield algebra which tells us how to interpret the possible word orders generated by a given tree. This “multiple transduction” approach has several other instances and though it is conceivable that such multiple transductions may help alleviate some of the *symptoms*, it does not address their root cause, which is the asymmetry of the system.

What we need is a system which can reject “candidate derivations” which have been computed at the previous level. Such a solution can be found in (Muskens 2007), who produces first-order logic formulas and adds a separate theorem-prover component (which is essentially a model-builder for multimodal categorial grammars). Muskens’ solution would solve the problems with lambda grammars by essentially generating potential derivations and asking a multimodal grammar if these derivations are valid. However, this setup does not seem to have any benefits over a direct multimodal implementation and suffers from the same complexity problems as multimodal categorial grammars. Pogodalla & Pompigne (2012), discussed below since they change the types in addition to the terms, also fall into this category.

Another potential solution in this family would be to add *term equations* (and corresponding reductions) to the lambda calculus. However, it is unclear what sort of form such a solution would take.

As we have seen in the treatment of transitive verb conjunctions, adding case offers (at best) a partial solution by attacking the symptoms rather than the underlying cause of the problem. It should also be noted that the only solution of this kind which has been worked out in any detail uses dependent types and this complicates the *types* as well as the terms (Pogodalla & Pompigne 2012, Pompigne 2013), which moves us to the next point.

3. Change the *types*. Various authors (de Groote & Maarek 2007, Pogodalla & Pompigne 2012) have looked at extending the type theory of lambda grammars beyond the simply typed lambda calculus. Of these extensions, dependent types seem well-suited to the challenges posed in this paper, though they would need to be added on a much larger scale than previously assumed and they would complicate the type/term calculus and its mathematical properties considerably.

An alternative solution, proposed in the context of linear grammars (Worth 2014) uses subtyping combined with restrictions on the form of subterm cooccurrences. It is unclear to me at the moment whether this type of treatment is equivalent to other proposals (eg. those of hybrid type-logical grammars) and whether it corresponds to a natural fragment of first-order linear logic.

It seems that the easiest way to fix the inadequacies of lambda grammars would be to extend the system to hybrid type-logical grammar: existing linguistic analyses in lambda grammars can be preserved and/or corrected while the system keeps much of the flavor of lambda grammars.

An alternative, especially for those convinced of the need to extend lambda grammars to handle linguistic features and island constraints (Pogodalla & Pompigne 2012, Pompigne 2013), is to move to first-order linear logic, which also allows us to preserve the things that work in lambda grammars but incorporate a simple treatment of both features and island constraints without having to change the underlying logical theory. Those particularly attached to dependent types can obtain them from MILL1 proofs by means of the Curry-Howard isomorphism; first-order logic is a fairly weak fragment of the lambda calculus with dependent types (Sørensen & Urzyczyn 2006) so we need to verify whether it is expressive enough. Since smart parsing algorithms for lambda grammars (de Groote 2007) already use first-order (linear) logic to drive proof search, this solution stays close to the computational core of lambda grammars: it remedies the severe problems but also allows us to include treatments for which much more complicated analyses have been proposed. However, it is not clear in such a setup what the typed lambda terms actually contribute and we would have a much simpler system if we simply removed the typed lambda terms from the surface structure component and handle all of the surface structure in first-order linear logic.

8 Conclusions

In this paper, we have shown that Hybrid Type-Logical Grammars (Kubota & Levine 2013a) (and by extension lambda grammars/abstract categorial grammars) can be embedded in first-order linear logic by means of a simple translation, formula to formula and proof to proof. This provides cleaner proof-theoretic foundations for Hybrid Type-Logical Grammars but also suggests new ways of parsing these grammars. As an immediate corollary, we have also shown that Hybrid Type-Logical Grammars are NP-complete (like lambda grammars and the Lambek calculus).

We have also seen how this translation provides a new perspective of the known (but often ignored) problems of lambda grammars with coordination and shown that the lack of left-right symmetry (or, at the very least, the absence of a way to emulate the Lambek calculus introduction rules) results in overgeneration and descriptive inadequacy problems for a much larger class of cases than previously assumed.

Combined with the results from (Moot & Piazza 2001) and (Moot 2014), this means that the Lambek calculus, the Displacement calculus, lambda grammars and Hybrid Type-Logical Grammars can all be translated into first-order linear logic by means of simple translations and that, moreover, many of the analyses of linguistic phenomena in these different systems converge upon translation into first-order linear logic.

First-order linear logic can thus be seen as a way to decompose the connectives of all these logics, separating the functor/argument structure from the word order operations.

Acknowledgments

This paper is deeply indebted to Yusuke Kubota and Robert Levine, whose ESSLLI 2013 course awoke my curiosity both about the proof theoretic aspects of hybrid type-logical grammar and about the descriptive inadequacies of lambda grammars/abstract categorial grammars — the two principal themes of the current paper.

Early versions of these ideas were presented at the LIX Colloquium on the Theory and Application of Formal Proofs (Palaiseau, November 2013), Computational Linguistics in the Netherlands (Leiden, January 2014) and the Polymnie workshop (Toulouse, March 2014). I would like all the people present there for their questions and constructive comments, notably Crit Cremers, Philippe de Groote, Dominic Hughes and Dale Miller.

Last, but certainly not least, I would like to thank Michael Moortgat, Carl Pollard and Christian Retoré for their discussion about the themes of this paper.

All remaining errors are of course my own.

This work has benefitted from the generous support of the French agency Agence Nationale de la Recherche as part of the project Polymnie (ANR-12-CORD-0004).

References

- Andreoli, J.-M. (1992), ‘Logic programming with focussing proofs in linear logic’, *Journal of Logic and Computation* **2**(3).
- Barendregt, H., Dekkers, W. & Statman, R. (2013), *Lambda Calculus with Types*, Perspectives in Logic, Cambridge University Press.
- Bayer, S. & Johnson, M. (1995), Features and agreement, in ‘Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics’, San Francisco, pp. 70–76.
- Bellin, G. & van de Wiele, J. (1995), Empires and kingdoms in MLL, in J.-Y. Girard, Y. Lafont & L. Regnier, eds, ‘Advances in Linear Logic’, Cambridge University Press, pp. 249–270.

- van Benthem, J. (1995), *Language in Action: Categories, Lambdas and Dynamic Logic*, MIT Press, Cambridge, Massachusetts.
- Bourreau, P. (2013), Traitements d’ellipses : deux approches par les grammaires catégorielles abstraites, *in* ‘Proceedings of Traitement Automatique des Langues Naturelles (TALN)’, Les Sables d’Olonne.
- Brock-Nannestad, T. & Schürmann, C. (2010), Focused natural deduction, *in* C. G. Fermüller & A. Voronkov, eds, ‘Logic for Programming, Artificial Intelligence, and Reasoning’, Vol. 6397 of *Lecture Notes in Computer Science*, Springer, pp. 157–171.
- Buszkowski, W. (1996), ‘Extending Lambek grammars to basic categorial grammars’, *Journal of Logic, Language and Information* **5**(3–4), 279–295.
- Curry, H. B. (1961), Some logical aspects of grammatical structure, *in* R. Jakobson, ed., ‘Structure of Language and its Mathematical Aspects, Proceedings of the Symposia in Applied Mathematics’, Vol. XII, American Mathematical Society, pp. 56–68.
- Danos, V. & Regnier, L. (1989), ‘The structure of multiplicatives’, *Archive for Mathematical Logic* **28**, 181–203.
- Dörre, J. & Manandhar, S. (1995), Constraint-based Lambek calculi, *in* P. Blackburn & M. de Rijke, eds, ‘Specifying Syntactic Structures. Studies in Logic, Language and Information’, CSLI, Stanford.
- Girard, J.-Y. (1987), ‘Linear logic’, *Theoretical Computer Science* **50**, 1–102.
- Girard, J.-Y. (1991), Quantifiers in linear logic II, *in* G. Corsi & G. Sambin, eds, ‘Nuovi problemi della logica e della filosofia della scienza’, Vol. II, CLUEB, Bologna, Italy. Proceedings of the conference with the same name, Viareggio, Italy, January 1990.
- Girard, J.-Y. (1995), Linear logic: Its syntax and semantics, *in* J.-Y. Girard, Y. Lafont & L. Regnier, eds, ‘Advances in Linear Logic’, Cambridge University Press, pp. 1–42.
- Girard, J.-Y., Lafont, Y. & Taylor, P. (1988), *Proofs and Types*, Cambridge Tracts in Theoretical Computer Science 7, Cambridge University Press.
- de Groote, P. (2001), Towards abstract categorial grammars, *in* ‘Proceedings of the 39th Annual Meeting on Association for Computational Linguistics’, Association for Computational Linguistics, pp. 252–259.
- de Groote, P. (2002), Tree adjoining grammars as abstract categorial grammars, *in* ‘TAG+6, Proceedings of the sixth International Workshop on Tree Adjoining Grammars and Related Frameworks’, Università di Venezia, pp. 145–150.

- de Groote, P. (2007), ‘Abstract categorial grammar parsing: the general case’, Talk given at the Colloquium in Honor of Gérard Huet, June 22-23, Paris.
- de Groote, P. & Maarek, S. (2007), Type-theoretic extensions of abstract categorial grammars, *in* R. Muskens, ed., ‘Proceedings of Workshop on New Directions in Type-theoretic Grammars’, pp. 19–30.
- Guerrini, S. (1999), Correctness of multiplicative proof nets is linear, *in* ‘Fourteenth Annual IEEE Symposium on Logic in Computer Science’, IEEE Computer Science Society, pp. 454–263.
- Hendriks, P. (1995), Ellipsis and multimodal categorial type logic, *in* G. Morrill & R. T. Oehrle, eds, ‘Proceedings of Formal Grammar 1995’, Barcelona, Spain, pp. 107–122.
- Bindley, J. R. (2008), *Basic Simple Type Theory*, Vol. 42 of *Cambridge Tracts in Theoretical Computer Science*, Cambridge University Press.
- Kanazawa, M. (2011), Parsing and generation as datalog query evaluation, Technical report, National Institute of Informatics, Tokyo.
- Kubota, Y. & Levine, R. (2012), Gapping as like-category coordination, *in* D. Béchet & A. Dikovsky, eds, ‘Logical Aspects of Computational Linguistics’, Vol. 7351 of *Lecture Notes in Computer Science*, Springer, Nantes, pp. 135–150.
- Kubota, Y. & Levine, R. (2013a), Coordination in hybrid type-logical grammar, *in* ‘Ohio State University Working Papers in Linguistics’, Vol. 60, Columbus, Ohio.
- Kubota, Y. & Levine, R. (2013b), Determiner gapping as higher-order discontinuous constituency, *in* G. Morrill & M.-J. Nederhof, eds, ‘Formal Grammar’, Vol. 8036 of *Lecture Notes in Computer Science*, Springer, pp. 225–241.
- Kubota, Y. & Levine, R. (2013c), Empirical foundations for hybrid type-logical categorial grammar, *in* ‘ESSLLI 2013 Course Notes’, Düsseldorf.
- Kurtonina, N. & Moortgat, M. (1997), Structural control, *in* P. Blackburn & M. de Rijke, eds, ‘Specifying Syntactic Structures’, CSLI, Stanford, pp. 75–113.
- Lamarche, F. & Retoré, C. (1996), Proof nets for the lambek calculus — an overview, *in* V. M. Abrusci & C. Casadio, eds, ‘Proofs and Linguistic Categories’, CLUEB, Bologna, pp. 241–262.
- Lambek, J. (1958), ‘The mathematics of sentence structure’, *American Mathematical Monthly* **65**, 154–170.
- Lincoln, P. & Shankar, N. (1994), Proof search in first-order linear logic and other cut-free sequent calculi, *in* ‘Proceedings of Logic in Computer Science (LICS’94)’, IEEE Computer Society Press, pp. 282–291.

- Matsuzaki, T., Miyao, Y. & Tsujii, J. (2007), Efficient HPSG parsing with supertagging and CFG-filtering, *in* ‘Proceedings of the 20th international joint conference on Artificial intelligence’, pp. 1671–1676.
- Moortgat, M. (1996*a*), Generalized quantifiers and discontinuous type constructors, *in* H. Bunt & A. van Horck, eds, ‘Discontinuous Constituency’, Mouton de Gruyter, Berlin, pp. 181–207.
- Moortgat, M. (1996*b*), ‘Multimodal linguistic inference’, *Journal of Logic, Language and Information* **5**(3–4), 349–385.
- Moortgat, M. (1997), Categorical type logics, *in* J. van Benthem & A. ter Meulen, eds, ‘Handbook of Logic and Language’, Elsevier/MIT Press, chapter 2, pp. 93–177.
- Moot, R. (2002), Proof Nets for Linguistic Analysis, PhD thesis, Utrecht Institute of Linguistics OTS, Utrecht University.
- Moot, R. (2007), Filtering axiom links for proof nets, *in* L. Kallmeyer, P. Monachesi, G. Penn & G. Satta, eds, ‘Proceedings of Formal Grammar 2007’. to appear with CSLI.
- Moot, R. (2014), Extended lambek calculi and first-order linear logic, *in* C. Casadio, B. Coecke, M. Moortgat & P. Scott, eds, ‘Categories and Types in Logic, Language, and Physics: Essays dedicated to Jim Lambek on the Occasion of this 90th Birthday’, number 8222 *in* ‘Lecture Notes in Artificial Intelligence’, Springer, pp. 297–330.
- Moot, R. & Piazza, M. (2001), ‘Linguistic applications of first order multiplicative linear logic’, *Journal of Logic, Language and Information* **10**(2), 211–232.
- Moot, R. & Retoré, C. (2012), *The Logic of Categorical Grammars: A Deductive Account of Natural Language Syntax and Semantics*, Lecture Notes in Artificial Intelligence, Springer.
- Morrill, G., Valentín, O. & Fadda, M. (2011), ‘The displacement calculus’, *Journal of Logic, Language and Information* **20**(1), 1–48.
- Murawski, A. S. & Ong, C.-H. L. (2000), Dominator trees and fast verification of proof nets, *in* ‘Logic in Computer Science’, pp. 181–191.
- Muskens, R. (2001), Categorical grammar and lexical-functional grammar, *in* ‘Proceedings of the LFG01 Conference’, University of Hong Kong, pp. 259–279.
- Muskens, R. (2003), Languages, lambdas and logic, *in* G.-J. Kruijff & R. T. Oehrle, eds, ‘Resource Sensitivity in Binding and Anaphora’, Studies in Linguistics and Philosophy, Kluwer, pp. 23–54.

- Muskens, R. (2007), ‘Separating syntax and combinatorics in categorial grammar’, *Research on Language and Computation* **5**(3), 267–285.
- Oehrle, R. T. (1994), ‘Term-labeled categorial type systems’, *Linguistics & Philosophy* **17**(6), 633–678.
- Pentus, M. (1995), Lambek grammars are context free, in ‘Proceedings of the Eighth Annual IEEE Symposium on Logic in Computer Science’, Montreal, Canada, pp. 429–433.
- Pentus, M. (1997), ‘Product-free Lambek calculus and context-free grammars’, *Journal of Symbolic Logic* **62**, 648–660.
- Pereira, F. & Shieber, S. (1987), *Prolog and Natural Language Analysis*, CSLI, Stanford.
- Pogodalla, S. & Pompigne, F. (2012), Controlling extraction in abstract categorial grammars, in P. de Groote & M.-J. Nederhof, eds, ‘Proceedings of Formal Grammar 2010–2011’, Vol. 7395 of *LNCS*, Springer, pp. 162–177.
- Pollard, C. (2011), Proof theoretic background for linear grammar, Technical report, Ohio State University.
- Pompigne, F. (2013), Modélisation Logique de la Langue et Grammaire Catégorielles Abstraites, PhD thesis, Université de Lorraine.
- Roorda, D. (1991), Resource Logics: A Proof-theoretical Study, PhD thesis, University of Amsterdam.
- Savateev, Y. (2009), Product-free lambek calculus is NP-complete, in ‘Symposium on Logical Foundations of Computer Science (LFCS) 2009’, pp. 380–394.
- Shieber, S., Schabes, Y. & Pereira, F. (1995), ‘Principles and implementation of deductive parsing’, *Journal of Logic Programming* **24**(1–2), 3–36.
- Sørensen, M. H. & Urzyczyn, P. (2006), *Lectures on the Curry-Howard Isomorphism*, Vol. 143 of *Studies in Logic and the Foundations of Mathematics*, Elsevier.
- Steedman, M. (2001), *The Syntactic Process*, MIT Press, Cambridge, Massachusetts.
- Troelstra, A. S. & Schwichtenberg, H. (2000), *Basic Proof Theory*, Vol. 43 of *Cambridge Tracts in Theoretical Computer Science*, 2 edn, Cambridge University Press.
- Worth, C. (2014), The phenogrammar of coordination, in ‘EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)’, pp. 28–36.